AF/JIM
1638

FULBRIGHT & JAWORSKI L.L.P.
A REGISTERED LIMITED LIABILITY PARTNERSHIP
600 CONGRESS AVENUE, SUITE 2400
AUSTIN, TEXAS 78701-3271
WWW.FULBRIGHT.COM

RHANSON@FULBRIGHT.COM
DIRECT DIAL: (512) 536-3085

TELEPHONE: (512) 474-5201
FACSIMILE: (512) 536-4598

May 24, 2004

CERTIFICATE OF MAILING
37 C.F.R. 1.8

I hereby certify that this correspondence is being deposited with the U.S. Postal Service with sufficient postage as First Class Mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date below:

May 24, 2004
Date

Robert E. Hanson

Mail Stop Appeal Brief-Patent
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Re: *SN 09/602,840 entitled "METHOD FOR ALTERING THE NUTRITIONAL CONTENT OF PLANT SEED" by Julie A. Kirihara, et al.*
Our Ref. DEKM:180USD1; Client Ref. 51209 US 23

Commissioner:

Transmitted herewith for filing are:

1. A Brief on Appeal (an original and two copies);

2. A check for $330.00 to cover the filing fee for the appeal brief; and

3. A return postcard to acknowledge receipt of these materials. Please date stamp and mail this postcard.

If the check is inadvertently omitted, or the amount is insufficient, or should any additional fees under 37 C.F.R. §§ 1.16 to 1.21 be required for any reason relating to the enclosed materials, or should an overpayment be included herein, the Commissioner is authorized to deduct or credit said fees from or to Fulbright & Jaworski L.L.P. Account No.: 50-1212/DEKM:180USD1.

Respectfully submitted,

Robert E. Hanson
Reg. No. 42,628

REH/vv
Enclosures
25417389.1

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:
Julie A. Kirihara, et al.

Serial No.: 09/602,840

Filed: June 23, 2000

For: METHOD FOR ALTERING THE NUTRITIONAL CONTENT OF PLANT SEED

Group Art Unit: 1638

Examiner: Stuart F. Baum

Atty. Dkt. No.: DEKM:180USD1

CERTIFICATE OF MAILING
37 C.F.R. §1.8

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as First Class Mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the date below:

May 24, 2004
Date

Robert E. Hanson

BRIEF ON APPEAL

05/27/2004 EFLORES 00000134 09602840
01 FC:1402 330.00 OP

25413056.1

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

TABLE OF CONTENTS

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Julie A. Kirihara, et al.	Group Art Unit: 1638
Serial No.: 09/602,840	Examiner: Stuart F. Baum
Filed: June 23, 2000	Atty. Dkt. No.: DEKM:180USD1
For: METHOD FOR ALTERING THE NUTRITIONAL CONTENT OF PLANT SEED	

BRIEF ON APPEAL

Mail Stop Appeal Brief-Patent
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Appellants hereby submit an original and two copies of this Appeal Brief in response to the Final Office Action dated December 22, 2003. The fee for filing this Appeal Brief is attached hereto. This Brief is filed pursuant to the Notice of Appeal mailed March 22, 2004. The date for filing the instant Brief is May 24, 2004, based on the receipt of the Notice of Appeal by the Patent and Trademark Office on March 24, 2004.

No additional fees are believed due in connection with the instant paper. However, should any other fees be due, or the attached fee be deficient or absent, the Commissioner is authorized to withdraw the appropriate fee from Fulbright & Jaworski L.L.P. Deposit Account No. 50-1212/DEKM:180USD1. Please date stamp and return the enclosed postcard to evidence receipt of this document.

I. REAL PARTIES IN INTEREST

The real party in interest is Monsanto Company, the parent of wholly-owned subsidiary DEKALB Genetics Corporation, the assignee of this application.

II. RELATED APPEALS AND INTERFERENCES

There are no interferences or appeals for related cases.

III. STATUS OF THE CLAIMS

Claims 1-100 were filed with the application on June 23, 2000. Claims 1-71, 74-77, 80-83, 85, 97 and 92-93 were canceled in a Preliminary Amendment filed concurrently with the application. Claims 100-114 were added in a Supplemental Preliminary Amendment, mailed in the case on November 1, 2001.

Claims 72-73, 78-79, 84, 86, 88-91, 94-110 were pending at the time of the final Office Action mailed in the case on September 27, 2002. After the filing of a Notice of Appeal and Appeal Brief by Appellants, the Examiner issued a Third Office Action mailed on September 27, 2002. In a Response to the Third Office Action, Appellants canceled claim 94.

A final Office Action was mailed by the Examiner on December 22, 2003. No amendments have been filed after the final Office Action. Claims 72-73, 78-79, 84, 86, 88-91, 95-110 are therefore now pending and are the subject of the instant appeal. A copy of the appealed claims is attached as APPENDIX 1.

IV. STATUS OF AMENDMENTS

No amendments were made subsequent to the final Office Action.

V. SUMMARY OF THE INVENTION

The invention relates to transgenic maize plants having increased starch content and extractability and methods of use thereof. Specification at page 6, line 13 to page 7, line 2. Increased starch content and extractability is obtained by the expression of RNA molecules that are substantially identical or complementary to 19kD or 22kD α-zein plant seed storage proteins, thereby decreasing the amount of corresponding seed storage protein and increase in starch content and extractability in the cells of the plant. Specification at page 6, line 13 to page 7, line 23.

VI. ISSUES ON APPEAL

(1) Are claims 72-73, 78-79, 84, 86, 88-91 and 94-110 properly rejected under 35 U.S.C. §112, second paragraph, as being indefinite?

(2) Are claims 72-73, 78-79, 84, 86, 88-91 and 94-110 properly rejected under 35 U.S.C. §112, first paragraph, as not being supported by an adequate written description in the specification?

(3) Are claims 72-73, 78-79, 84, 86, 88-91 and 94-110 properly rejected under 35 U.S.C. §112, first paragraph, as not being enabled by the specification?

VII. GROUPING OF THE CLAIMS

The claims stand or fall together.

VIII. SUMMARY OF THE ARGUMENT

A. The terms "augmented," "preselected," "substantially identical" and "substantially complementary" are not indefinite under 35 U.S.C. § 112, second paragraph, as they are used in the claims, as one of skill in the art would readily understand the meaning of the terms when viewed in light of the language of the entire claim, the teaching of the specification and the knowledge of those of skill in the art.

B. The claims do not lack written description because they are directed to a discrete class of subject matter that is fully described in the specification. The specification describes the structural characteristics of 19 kD or 22 kD α-zein plant seed storage proteins by exemplary coding sequences and conserved ***functional domains***. In addition, more than ***70 genes encoding zein protein were isolated prior to the filing of the application.*** The structural detail provided is well more than that required under §112. The claims therefore do not lack written description.

C. The Examiner has failed to establish a *prima facie* case of lack of enablement under 35 U.S.C. § 112, first paragraph, and improperly ignores evidence submitted by Appellant affirmatively demonstrating the enablement of the claims. The alleged evidence presented in support of the rejection in fact supports the enablement of the claims and contradicts the rejections made.

VIII. ARGUMENT

A. The Claims Are Not Indefinite Under 35 U.S.C. §112, Second Paragraph

The Examiner has finally rejected claims 72-73, 78-79, 84, 86, 88-91 and 94-110 under 35 U.S.C. §112, second paragraph, as being indefinite for failing to particularly point out the

subject matter which Applicant regards as the invention. The rejections and Appellants' response thereto are set forth below.

(1) Rejection of Claim 72

(a) "Augmented"

The Examiner asserts that the claim term "augmented" is unclear in that it has not been defined and it is allegedly not clear whether this indicates incorporation of a nucleic acid into the genome. However, the meaning of the term is clear from the language of the claim. Claim terms must be viewed in the context of the claim in which they are found. In claim 72, a fertile transgenic *Zea mays* plant is recited "the ***genome*** of which is ***stably augmented*** by a preselected DNA sequence." (emphasis added). The claim further specifies that "said preselected DNA sequence is ***transmitted*** through a ***complete normal sexual cycle*** of the transgenic plant to the ***next generation***." (emphasis added). In view of the claim language, the only reasonable interpretation is that augmented refers to genetic transformation in the genome.

Were the foregoing not the case the term "genome" in claim 72 would be completely superfluous. Thus interpretations excluding involvement of the genome may not be made as was done by the Examiner. The claim further specifies that the preselected DNA is transmitted through a normal sexual cycle and that the genome is "stably augmented." This also indicates stable transformation in the genome.

The Examiner has therefore failed to give the claims a fair and reasonable reading based on the language of the claim as whole. Properly viewed, the term is fully definite. Removal of the rejection is therefore respectfully requested.

(b) "Preselected"

The Examiner rejects the use of the term "preselected" as indefinite because it has not been defined. In response, Appellants note that the term has a well known meaning in the art. Appellants direct the Examiner to the online version of the Merriam Webster® dictionary (http://www.m-w.com/cgi-bin/dictionary), which gives the definition of the present tense of preselected, "preselect", as being "to choose in advance usually on the basis of a particular criterion." **(See Exhibit A)** The dictionary meaning demonstrates the well known meaning of the term. There is nothing indefinite in the use of a term having a well known meaning in the art.

In the final Office Action maintaining the rejection the Examiner further asserted that the term was indefinite because, while acknowledging that "the Office understand the meaning of the word 'preselected'", the criteria by which the sequence is selected have allegedly not been defined. In response, it is noted that this is irrelevant. The only criteria that is necessary is that the DNA be selected for use in the claim. One of skill in the art may use any criteria in selecting the DNA for use in the claim. However, this does not render the claim indefinite because breadth does not equate to indefiniteness, as apparently believed here by the Examiner. The Action itself acknowledges that "preselected" has a known meaning. This alone establishes that one of skill in the art would understand the meaning of the claim and thus that the claim is definite. What criteria are used to preselect are irrelevant because they are not a part of the claim and there is no basis in law to attempt to make Appellants add such a criteria given that the full metes and bounds of the claim are clear. Given that the scope of the claim is clear to those of skill in the art, the claim is fully definite in compliance with the second paragraph of §112.

Removal of the rejection is thus respectfully requested.

(2) Rejections of claims 72, 88 and 90

Claims 72, 88 and 90 were finally rejected as allegedly being indefinite for the recitation of "substantially identical" and "substantially complementary to all or a portion." In particular, it is stated that the recited phrases are vague and unclear and do not specify what portion or percent of the sequence applicants are referring to.

In response, it is noted that the terms "substantially identical" and "substantially complementary" are defined in the specification at page 12, lines 11-24. The use of the terms in the claims is thus not indefinite. Additionally, the terms are further defined by the language of claim 72. For example, claim 72, from which claims 88 and 90 depend, reads as follows:

> A fertile transgenic *Zea mays* plant having an increased starch content, the genome of which is stably augmented by a preselected DNA sequence encoding an RNA molecule which is ***substantially identical, or complementary***, to an mRNA encoding a 19kD or a 22kD α-zein plant seed storage protein, wherein the preselected DNA sequence is ***expressed*** in the cells of the transgenic plant in an amount ***sufficient to decrease*** the amount of said seed storage protein and increase starch content in the cells of a plant which only differ from the cells of said transgenic plant in that said preselected DNA sequence is absent, and wherein said preselected DNA sequence is transmitted through a complete normal sexual cycle of the transgenic plant to the next generation.

(emphasis added)

As can be seen, the clam indicates that expression of the preselected DNA sequence results in a decrease of the amount of seed storage protein and increased starch content in the cells of a plant. Claims 88 and 90 incorporate this limitation by dependency. Thus the meaning of the terms to one of skill in the art is clear, *e.g.*, that substantially identical, or complementary ***must refer*** to a preselected DNA sequence encoding an RNA molecule that is sufficiently identical or complementary to an mRNA encoding a 19kD or a 22kD α-zein plant seed storage protein to ***reduce expression*** of a 19kD or 22kD α-zein seed storage protein. This is so because the claim specifies that the preselected DNA sequence is expressed in the cells of the transgenic plant in an amount sufficient to decrease the amount of seed storage protein and increase starch content in the cells of a plant comprising the cell. Such a sequence could not represent a single

homologous base pair as alleged by the Examiner, as it is well known to those of skill in the art that in order to form the type of stable complex required for antisense suppression, longer stretches of complementary sequences are required.

One of skill in the art would thus fully understand the metes and bounds of the claim because the claim is defined by discrete and readily ascertainable criteria. The test for definiteness under 35 U.S.C. § 112, second paragraph, is whether "those skilled in the art would understand what is claimed when the claim is read in light of the specification." *Orthokinetics, Inc. v. Safety Travel Chairs, Inc.*, 806 F.2d 1565, 1576, 1 USPQ2d 1081, 1088 (Fed. Cir. 1986). If one skilled in the art is able to ascertain the meaning of the claim, 35 U.S.C. § 112, second paragraph, is satisfied. *Id.* In view of the definition in the specification and claim language, the referenced term fully meet this requirement.

In conclusion, Appellants note that the rejected claims are fully definite in compliance with 35 U.S.C. § 112, second paragraph. Reversal of the rejection is thus respectfully requested.

B. The Claims Do Not Lack Written Description Under 35 U.S.C. §112, First Paragraph

The Examiner added a rejection of claims 72-73, 78-79, 84, 86, 88-91, and 94-110 under the first paragraph of 35 U.S.C. §112 in the Third Office Action based on an alleged lack of written description for the claimed invention. In particular, it was alleged that the specification does not describe the structural characteristics of 19 kD or 22 kD α-zein plant seed storage proteins.

The rejection is utterly without merit and its reissuance and maintenance by the Examiner is puzzling to Appellants. Essentially the same rejection was issued in the first Office Action in

this case, but was withdrawn in the second Office Action in view of the persuasive comments of Appellants in the Response to the First Office Action. Overwhelming evidence supporting the structural characteristics of 19 kD or 22 kD α-zein genes was shown at that time. The Third Office Action nonetheless reissued the rejection and the Examiner has now made the rejection final.

The Examiner has alleged that written description is lacking in particular because information was not provided regarding what sequences are substantially identical or complementary to an mRNA encoding 19 kD or 22 kD α-zein genes. However, as set forth above, the claim terms "substantially identical, or complementary" define a discrete class of subject matter that is well more than adequately described. As described, such sequences must hybridize *in vivo* to 19 kD or 22 kD α-zein genes so as to yield a decrease of expression of the 19 kD or 22 kD α-zein genes. This subject matter is fully defined in the specification as set forth below.

It is first noted that Appellants do not claim 19 kD and 22 kD α-zein plant seed storage protein genes *per se*, as this class of genes was ***known at the time the application was filed.*** Rather transgenic plants expressing these proteins or methods comprising the use thereof are claimed. The structural features unique to a maize 19 kD and 22 kD α-zein plant seed storage protein have been fully described in the specification. Accordingly, it can not be said that Appellants lack written description for the terms. For example, the Board's attention is directed to FIG. 1 of the application. There shown are ***functional domains*** that are ***conserved*** and shared among the zeins. Further attention is drawn to pages 1-3 of the application. There, the specification describes the family of ***known zeins***, including 19 kD and 22 kD α-zeins. At page 2, it is indicated, with a citation to Rubenstein (1982), that ***over 70 genes encoding zein protein***

have been isolated. Further on page 2, at lines 19-24, ***functional domains of 19 kD α-zeins are described.*** The structural characteristics of the 19 kD and 22 kD α-zein plant seed storage proteins have thus been described in full compliance with 35 U.S.C. §112, first paragraph and *Eli Lilly* is therefore inapplicable to the instant situation.

The specification still further describes 19 kD and 22 kD α-zein seed storage proteins in the form of the ***nucleic acid sequence of the A20 and Z4 cDNAs***, respectively. While these sequences are species of 19 kD and 22 kD α-zein protein genes, the species are representative of the genus as evidenced by Marks *et al.* (1985) **(Exhibit B)**, which was cited previously by the Examiner in an enablement rejection. Marks *et al.* demonstrates the ***common structural characteristics shared among 19 kD and 22 kD α-zeins.*** For example, in the first sentence of the Abstract of Marks *et al.*, it is indicated that a comparison of the protein and DNA sequences of zein cDNA clones "reveals that they share extensive sequence homology and probably originated from a common ancestral gene." In the first paragraph of the Discussion section it is indicated that cDNA sequences among the 19 kD and 22 kD group of α-zein sequences are ***75 to 95% and 92% homologous***, respectively. Further, Marks *et al.* provides ***sequence information and comparisons*** among 19 kD and 22 kD α-zeins. The disclosure of Marks *et al.* thereby demonstrates the shared structural characteristics of the 19 kD and 22 kD α-zein seed storage proteins. Combined with Appellants' disclosure of the structural characteristics of 19 kD and 22 kD α-zein proteins, this is more than adequate to demonstrate compliance with the written description requirement.

In view of the foregoing, Appellants assert that the written description requirement has been fully satisfied. Reversal of the rejection under 35 U.S.C. §112, first paragraph is thus respectfully requested.

C. The Claims Are Enabled Under 35 U.S.C. §112, First Paragraph

Claims 72-73, 78-79, 84, 86, 88-91 and 95-110 were finally rejected under the first paragraph of 35 U.S.C. §112 as allegedly not being enabled by the specification. In particular, the Examiner alleged that the specification does not teach one of skill in the art how to increase the starch content or starch extractability or kernel hardness of seeds. It is stated that the specification only teaches how to make maize seeds with decreased amounts of the amino acid leucine and increased lysine by transforming *Zea mays* plants with SEQ ID NOs:1 and 2 operably linked to a Z10 promoter.

Appellants submit that no basis has been provided to doubt the enablement of the claims and further that enablement has been affirmatively demonstrated. The Examiner attempts to support the rejection by citing Coleman *et al.* (1997) (**Exhibit C**). Although the Examiner dropped the reference of Marks *et al.* (1985) (**Exhibit B**), this was cited in the enablement rejection in the first Office Action. As set forth below, both of these references affirmatively ***demonstrate*** the enablement of the invention.

Coleman *et al.* was cited as showing that high-lysine mutants exhibiting a reduction of α-zein content were "concomitant with an inferior endosperm quality." It was thus suggested that, because of the supposedly inferior endosperm quality, "reducing the α-zein protein content of maize seeds using the strategy of Appellants, will not increase the starch content or starch extractability of maize seeds." However, the reference does the very opposite and demonstrates the enablement of the claims. This "inferior" endosperm is in fact a ***soft and starchy endosperm***. (See Coleman, p. 7094, paragraph 2). The reference therefore shows the direct correlation between increased lysine, decreased α-zein and soft and starchy endosperm. The first Office

Action acknowledged this on page 5.[1] Demonstration of this correlation supports enablement, given that the Examiner already acknowledged that Appellants have demonstrated enablement for increasing lysine during the prosecution. The fact that Coleman *et al.* also demonstrated successful expression of a 24 kDa α-zein gene to induce the mutant phenotype provides still further evidence of enablement of the instant claims.

It is additionally noted that whether the endosperm is subjectively "inferior" or not is irrelevant to enablement. What is considered inferior for a plant used for one purpose, for example, for human consumption, is not necessarily the same as for a plant used for production of corn starch. Further, an "inferior" endosperm does not equate to an inability to increase starch content or extractability, as described in Coleman *et al.*

The other reference cited in the first Office Action, but apparently dropped for the Third and Final Office Actions, also demonstrates enablement. Marks *et al.* (1985) (**Exhibit B**), was initially cited as indicating that there are many different forms of 19 kD and 22 kD α-zeins with divergent and unpredictable functions. However, this reference ***supports enablement*** of the claims. Marks *et al.* states in the abstract that:

> A comparison of the DNA and protein sequences of a group of zein cDNA clones reveals that they ***share extensive sequence homology and probably originated from a common ancestral gene.*** A comparison of clones corresponding to Mr 22,000 polypeptides shows that they are 92% homologous, while five clones corresponding to Mr 19,000 zein vary in homology from 75 to 95%.
>
> (emphasis added)

Marks *et al.*, therefore, demonstrates the common structural characteristics shared among 19 kD and 22 kD α-zeins. In the first paragraph of the Discussion section of Marks *et al.*, it is indicated that cDNA sequences among the 19 kD and 22 kD group of α-zein sequences are 75 to 95% and

[1] See middle paragraph: "Two 'high –lysine' mutants were identified, opaque2 (o2) and floury2 (fl2)

92% homologous, respectively. Further, Marks *et al.* provides sequence information and comparisons among 19 kD and 22 kD α-zeins. The disclosure of Marks *et al.* thereby demonstrates the shared structural characteristics of the 19 kD and 22 kD α-zein seed storage proteins.

No basis has been provided by the Examiner to indicate that different isoforms have "divergent functions." The high degree of homology and indicated common ancestor among zeins strongly contradicts this. Additionally, there is no support in the Marks *et al.* reference for the prior contention that any of the mRNA isoforms described encode proteins other than zeins, the function of which is that of a seed storage protein. The reference, therefore, indicates the shared original and conserved structure of the 19 kD and 22 kD α-zein plant seed storage proteins.

The rejection adds citation to the reference of Moonan (2002) **(Exhibit D)**. The reference is alleged to show that sugarcane viral protection using less than 100% sequence identity resulted in "inferior viral protection." Appellants note however that this is irrelevant to enablement of the current claims. First, this was sugarcane, ***not maize*** as in the instant claims, thus there is no showing how this is relevant to the instant claims. Second, what is ***subjectively*** "inferior" has absolutely nothing to do with enablement. Enablement concerns teaching one of skill in the art how to make and use the claimed invention, not make and use something that is subjectively superior. Indeed, that statement cited in the Action demonstrates that viral protection ***was achieved***, but was apparently somehow subjectively inferior. Again, as long as one of skill in the art can make and use the invention, it is irrelevant if an invention is "inferior" to other techniques that may be available. Finally, the cited reference has to do with viral capsid

---

which have a higher lysine content due to a reduction in the α-zein protein content of the endosperm."

expression to obtain viral resistance. Again, there is absolutely no basis provided to analogize this to the current case. The current claims involve alteration of function of an ***endogenous maize gene***, whereas viral suppression has to do with infection from an ***external pathogen***. The scientific principles are entirely different than for the claimed invention. Appellants do not claim viral resistance. No basis has therefore been provided to draw any analogies between these technologies.

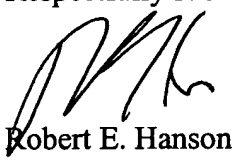
Further evidence of enablement of the claims is provided at page 83 of the specification. There it is described that endosperm cells in a maize kernel are made up primarily of large starch granules and protein sequestered in protein bodies. It is shown that a reduction in the number of protein bodies in endosperm cells derived from a transformant produced using a zein antisense construct was achieved. This was shown by light microscopy and the results given in FIG. 9. As can be seen in the figure, the results demonstrated a decrease in the amount of seed storage protein and therefore an increase in the relative starch content of the kernel.

In conclusion, the Examiner has failed to set forth any basis for doubting the enablement of the claims. Appellants have further affirmatively set forth evidence of the enablement of the claims that has not been contradicted by the Examiner. Reversal of the rejection is thus respectfully requested.

IX. CONCLUSION

It is respectfully submitted, in light of the above, that none of the appealed claims are properly rejected. Therefore, Appellants respectfully request that the Board reverse the pending grounds for rejection.

Respectfully submitted,



Robert E. Hanson
Reg. No. 42,628
Attorney for Appellants

FULBRIGHT & JAWORSKI L.L.P
600 Congress Avenue, Suite 2400
Austin, Texas 78701
(512) 536-3085

Date: May 24, 2004

APPENDIX 1: APPEALED CLAIMS

72. A fertile transgenic *Zea mays* plant having an increased starch content, the genome of which is stably augmented by a preselected DNA sequence encoding an RNA molecule which is substantially identical, or complementary, to an mRNA encoding a 19kD or a 22kD α-zein plant seed storage protein, wherein the preselected DNA sequence is expressed in the cells of the transgenic plant in an amount sufficient to decrease the amount of said seed storage protein and increase starch content in the cells of a plant which only differ from the cells of said transgenic plant in that said preselected DNA sequence is absent, and wherein said preselected DNA sequence is transmitted through a complete normal sexual cycle of the transgenic plant to the next generation.

73. A fertile transgenic *Zea mays* plant, the seeds of which have an increased starch extractability, the genome of said plant which is stably augmented by a preselected DNA sequence encoding an RNA molecule which is substantially identical, or complementary, to an mRNA encoding a 19kD or a 22kD α-zein plant seed storage protein, wherein the preselected DNA sequence is expressed in the seeds of the transgenic plant in an amount sufficient to decrease the amount of said seed storage protein and increase the starch extractability of the seed relative to the amount of said seed storage protein and starch extractability in the seeds of a plant which only differ from the seeds of said transgenic plant in that said preselected DNA sequence is absent, and wherein said preselected DNA sequence is transmitted through a complete normal sexual cycle of the transgenic plant to the next generation.

78. A seed derived from the plant of claim 72 or 73, wherein the seed comprises said preselected DNA sequence.

79. A progeny plant derived from the seed of claim 78, wherein the plant comprises said preselected DNA sequence.

84. The transgenic plant of claim 72 or 73, wherein the promoter comprises the 10 kD zein promoter.

86. The transgenic plant of claim 72 or 73, wherein the promoter comprises the 27kD zein promoter.

88. The transgenic plant of claim 72 or 73, wherein the preselected DNA sequence, which encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a seed storage protein, encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding 19 kD α-zein protein.

89. The transgenic plant of claim 72 or 73, wherein the preselected DNA sequence, which encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a seed storage protein, encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a 22 kD α-zein protein.

90. The transgenic plant of claim 72 or 73, wherein the preselected DNA sequence, which encodes an RNA molecule, substantially identical to all or a portion of an mRNA encoding a seed storage protein, encodes an RNA molecule substantially identical to all or a portion of an mRNA encoding a 19 kD α-zein protein.

91. The transgenic plant of claim 72 or 73, wherein the preselected DNA sequence, which encodes an RNA molecule substantially identical to all or a portion of an mRNA encoding a seed storage protein, encodes an RNA molecule substantially identical to all or a portion of an mRNA encoding a 22 kD α-zein protein.

95. The transgenic plant of claim 72 or 73, wherein the cell is transformed by a method selected from the group consisting of electroporation, microinjection, microprojectile bombardment, and liposomal encapsulation.

96. The transgenic plant of claim 78 or 79, further comprising stably transforming the cells with at least one selectable marker gene.

97. A method of producing a *Zea mays* seed with an increased starch content, comprising:

(a) growing a transgenic *Zea mays* plant, the genome of which is augmented with a preselected DNA sequence encoding an RNA molecule which is substantially identical, or complementary to an mRNA encoding a 19kD or a 22kD α-zein seed storage protein, wherein the preselected DNA sequence is expressed in the cells of the *Zea mays* plant in an amount sufficient to decrease the amount of seed storage protein; and

(b) selecting a seed from the transgenic *Zea mays* plant, wherein the seed has an increased amount of starch relative to the amount of starch in a seed selected from a plant which does not comprise said preselected DNA sequence.

98. A method of obtaining starch from a *Zea mays* seed, comprising:

(a) growing a transgenic *Zea mays* plant, the genome of which is augmented with a preselected DNA sequence encoding an RNA molecule which is substantially identical, or complementary, to an mRNA encoding a 19kD or a 22kD α-zein seed storage protein, wherein the preselected DNA sequence is expressed in the cells of the *Zea mays* plant in an amount sufficient to decrease the amount of seed storage protein;

(b) obtaining seed from said plaint; and

(c) extracting starch from the seed.

99. The method of claim 97 or 98 wherein the preselected DNA sequence is operably linked to a promoter functional in plant cells.

100. The method of claim 99 wherein the promoter comprises the 10 kD zein promoter.

101. The method of claim 99 wherein the promoter comprises the 27 kD zein promoter.

102. The method of claim 97 or 98 wherein the preselected DNA sequence encodes an RNA molecule which is substantially identical to all or a portion of the mRNA encoding a seed storage protein.

103. The method of claim 97 or 98 wherein the preselected DNA sequence encodes an RNA molecule which is substantially complementary to all or a portion of the mRNA encoding a seed storage protein.

104. The method of claim 102 wherein the preselected DNA sequence encodes an RNA molecule substantially identical to all or a portion of mRNA encoding a 19 kD α-zein protein.

105. The method of claim 102 wherein the preselected DNA sequence encodes an RNA molecule substantially identical to all or a portion of an mRNA encoding a 22 kD α-zein protein.

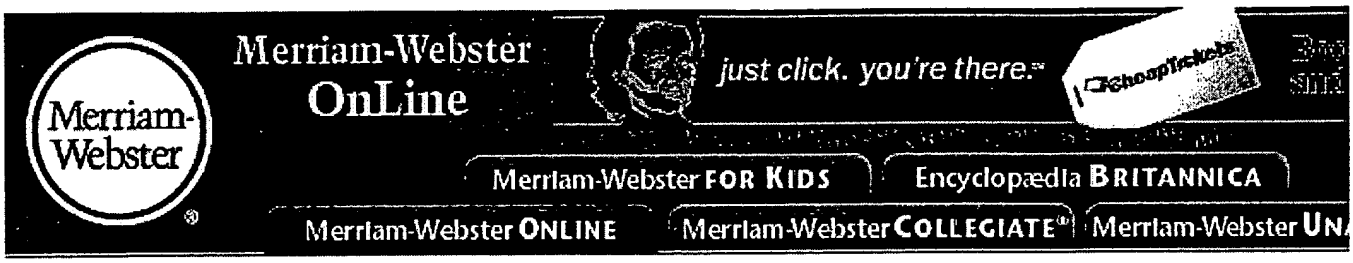
106. The method of claim 103 wherein the preselected DNA sequence encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a 19 kD α-zein protein.

107. The method of claim 103 wherein the preselected DNA sequence encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a 22 kD α-zein protein.

108. The method of claim 97 or 98 wherein the genome of the transgenic *Zea mays* plant is further augmented with a DNA sequence encoding a polypeptide that provides the transgenic *Zea mays* plant with increased kernel hardness.

109. The method of claim 97 or 98 wherein the transgenic *Zea mays* plant is produced from cells transformed by a method selected from the group consisting of electroporation, microinjection, microprojectile bombardment, and liposomal encapsulation.

110. The method of claim 97 or 98 wherein the genome of the transgenic *Zea mays* plant is further augmented with at least one selectable marker gene.



HOME
PREMIUM SERVICES
M-WCollegiate.com
M-WUnabridged.com
Britannica.com
Multi-User Licenses
DOWNLOADS
WORD OF THE DAY
WORD GAMES
WORD FOR THE WISE
ONLINE STORE
HELP
Merriam-Webster Inc.
Company information

Merriam-Webster Online Dictionary

Thesaurus

One entry found for **preselect**.

Main Entry: **pre·se·lect**
Pronunciation: "prE-s&-'lekt
Function: *transitive verb*
: to choose in advance usually on the basis of a particular criterion
- **pre·se·lec·tion** /-'lek-sh&n/ *noun*

For **More Information on "preselect" go to Britannica.com**

Get the **Top 10 Search Results for "preselect"**

Pronunciation Symbols

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1985 by The American Society of Biological Chemists, Inc.

Vol. 260, No. 30, Issue of December 25, pp. 16451-16459, 1985
Printed in U.S.A.

Nucleotide Sequence Analysis of Zein mRNAs from Maize Endosperm*

(Received for publication, May 10, 1985)

M. David Marks, Judith S. Lindell, and Brian A. Larkins‡

From the Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907

A comparison of the DNA and protein sequences of a group of zein cDNA clones reveals that they share extensive sequence homology and probably originated from a common ancestral gene. A comparison of clones corresponding to M_r 22,000 polypeptides shows they are 92% homologous, while five clones corresponding to the M_r 19,000 zeins vary in homology from 75 to 95%. The clones corresponding to the M_r 22,000 proteins are 60–65% homologous to clones encoding the M_r 19,000 zein proteins. A clone corresponding to the M_r 15,000 zein has little homology to either the M_r 22,000 or 19,000 zeins. Clones corresponding to both the M_r 22,000 and 19,000 zeins have two putative polyadenylation signals. S1 nuclease mapping indicates that the first polyadenylation signal following the stop codon is utilized by the M_r 22,000 sequences, while primarily the second polyadenylation signal is utilized by the M_r 19,000 sequences.

To study the expression of zein genes during maize endosperm development, we constructed and characterized a number of full-length cDNA clones (Marks *et al.*, 1985). On the basis of cross-hybridization studies, these clones were divided into nine distinct groups: one for the M_r 15,000 zein, five for the M_r 19,000 zeins, and three for the M_r 22,000 zeins. No clones for the M_r 10,000 or reduced soluble protein were isolated. However, the clones that were identified appear to represent a large proportion of the zein mRNA sequences in the endosperm mRNA population (Marks *et al.*, 1985).

Sequences of clones related to some of those in the groups we describe have been reported (Heidecker and Messing, 1983; Geraghty *et al.*, 1982; Spena *et al.*, 1982). Previous studies revealed that genes encoding the M_r 19,000 and 22,000 zein groups are homologous and probably have a common ancestral sequence (Marks and Larkins, 1982; Spena *et al.*, 1982), but the M_r 15,000 zein sequence was found to belong to an unrelated gene family.[1] Genes within the M_r 19,000 and 22,000 zein groups have diverged from one another through base substitutions, small insertions/deletions, and relatively large internal duplications (Heidecker and Messing, 1983; Marks and Larkins, 1982). Many of these genes contain two functional polyadenylation signals (Messing *et al.*, 1983), and several of the genes appear to have multiple transcriptional start sites (Langride and Feix, 1983).

This study presents a comprehensive comparison of sequences belonging to eight M_r 19,000 and 22,000 zein groups, and it presents the sequence of a full-length cDNA clone for a M_r 15,000 zein that contains a very long 5′ noncoding region. We have also analyzed the usage of the polyadenylation signals for the M_r 19,000 zein genes.

MATERIALS AND METHODS

DNA restriction endonucleases, Klenow fragment, T4 kinase, and S1 nuclease were purchased from Bethesda Research Laboratories. [γ-^{32}P]ATP, [α-^{32}P]dATP, and [α-^{32}P]dCTP were purchased from New England Nuclear or Amersham Corp. Nitrocellulose was obtained from Schleicher and Schuell. Calf intestinal alkaline phosphatase was purchased from Boehringer Mannheim.

5′-Terminal End Labeling and DNA Sequencing—Restriction fragments were prepared for sequencing by the radioactive end-labeling method of Maxam and Gilbert (1980) for 5′ ends as described by Pedersen *et al.* (1982). DNA fragments were sequenced by the method of Maxam and Gilbert (1980) with the modification of Krayev *et al.* (1980).

Computer Analysis—Sequences were analyzed using the Map and Gap programs developed by Devereux *et al.* (1984).

S1 Nuclease Mapping—Fragments for S1 nuclease mapping were either 5′ end labeled as described by Maxam and Gilbert (1980) or 3′ end labeled by filling in recessed 3′ ends with Klenow fragment as described by Maniatis *et al.* (1982). End-labeled fragments were cleaved by restriction enzyme digestion and resolved on a 5% polyacrylamide gel. The fragments of interest were excised from the gel and isolated by the crush-soak method of Maxam and Gilbert (1980).

S1 nuclease mapping was performed by the method of Berk and Sharp (1977) as described by Maniatis *et al.* (1982). One μg of endosperm poly(A) RNA plus 200 μg of tRNA were added to 3×10^4 to 1×10^5 cpm of labeled fragment in 20 μl of 40 mM Pipes[2] (pH 6.4), 1 mM EDTA (pH 8.0), 0.4 M NaCl, and 80% formamide. Reactions were heated to 80 °C for 10 min and then incubated overnight at 58 °C for Z15 fragments (% G + C = 55%) and at 52 °C for Z19 fragments (% G + C = 48%). The hybridization reaction was stopped by adding 0.38 ml of cold S1 nuclease buffer (0.28 M NaCl, 0.05 M NaOAc (pH 4.6), 4.5 mM $ZnSO_4$, and 20 μg/ml denatured calf thymus DNA) containing 400 units of S1 nuclease. The S1 digestion reactions were incubated at 37 °C for 30 min after which they were extracted with phenol/chloroform and the DNAs precipitated with ethanol. Samples were dissolved in sequencing dye (80% formamide, 0.5 × TNE (10 mM Tris-HCl, pH 7.4, 15 mM NaCl, and 1 mM EDTA), 0.1% bromphenol blue, and 0.1% xylene cyanol) and loaded onto a 6% polyacrylamide sequencing gel that contained 50% urea.

Northern Blot Analysis—Four μg of poly(A) RNA from membrane-bound polysomes of 18 day-after-pollination endosperms was resolved by electrophoresis on an agarose gel that contained methyl mercury hydroxide (1.4% agarose, 5 mM methyl mercury hydroxide, 50 mM boric acid, 5 mM sodium borate, 10 mM sodium sulfate, and 0.01 mM Na_2EDTA) (Bailey and Davidson, 1976). The gel was soaked in 0.5 M ammonium acetate and then stained with ethidium bromide. Gel slices that contained the upper and lower zein mRNA bands were excised and recast in separate 1.4% agarose gels. The RNA was transferred to nitrocellulose filters (Thomas, 1980), and the nitrocellulose was prehybridized for 8 h at 42 °C in 50% formamide, 5 × SSC (1 × SSC = 0.15 M NaCl, 0.015 M sodium citrate), 0.05 M sodium phosphate buffer, pH 6.8, 0.1% bovine serum albumin, 0.1% Ficoll,

* This is Journal Paper 10,319 of the Purdue Agricultural Experiment Station. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed.

[1] K. Pedersen, P. Argos, S. V. L. Nayarana, and B. A. Larkins, manuscript in preparation.

[2] The abbreviation used is: Pipes, 1,4-piperazinediethanesulfonic acid.

0.1% polyvinylpyrrolidone, 200 μg/ml heat-denatured sheared calf thymus DNA. The nitrocellulose was then hybridized overnight at 42 °C in 50% formamide, 5 × SSC, 0.05 M sodium phosphate buffer, pH 6.8, 0.1% bovine serum albumin, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 200 μg/ml sheared calf thymus DNA, 5% dextran sulfate, and 7×10^6 cpm of the isolated zein cDNA inserts labeled with ^{32}P by nick translation (Maniatis *et al.*, 1976). Following hybridization the nitrocellulose was washed twice in 1 × SSC containing 0.1% sodium dodecyl sulfate at room temperature and twice in 1 × SSC containing 0.1% sodium dodecyl sulfate at 68 °C. Filters were air dried and autoradiographed.

RESULTS

By cross-hybridization analysis at a stringent criterion (T_m −15 °C), we were able to distinguish nine groups of zein cDNA clones (Marks *et al.*, 1985). The DNA sequences of several members from three of these groups was previously reported (Marks and Larkins, 1982; Pedersen *et al.*, 1982).[1] In this study seven cDNA clones from the remaining six groups were completely sequenced by the method of Maxam and Gilbert (Fig. 1).

Analysis of M_r 22,000 Zein Sequences—Each of the clones, cZ22A-1 and cZ22B-1, encodes a complete zein protein, whereas the third, cZ22C-2, lacks a few coding bases on the 5′ end (Fig. 2). Both cZ22B-1 and cZ22C-2 contain an entire 3′ noncoding region, and cZ22A-1 contains a long 5′ noncoding region. The cZ22B-1 and cZ22C-2 sequences align without gaps; however, when these clones were compared to cZ22A-1, four short gaps were introduced to maximize homology. The largest gap results in a 9-base pair deletion in cZ22A-1 relative to cZ22B-1 and cZ22C-2. Overall, the deletion/insertion events have left cZ22B-1 and cZ22C-2 with three more amino acids than cZ22A-1. The clones differ from one another by 6–7%, and these nucleotide differences are fairly evenly distributed throughout the sequences. The longest stretch of perfect homology occurs in the region of the stop codon and is only 56 nucleotides in length. Forty-four per cent of the base changes have resulted in amino acid substitutions. In general, the physical properties of the substituted amino acids are conserved. For example, seven of the 26 amino acid changes that occurred during the divergence of cZ22A-1 and cZ22B-1 involved exchanges between valine and alanine (Fig. 3). A few substitutions have not been conservative. For example, at position 103 (Fig. 3) alanine is replaced with aspartic acid, and at positions 177 and 227 glutamic acid is replaced with glutamine and alanine, respectively. One other substitution resulted in a tryptophan residue at position 258. This amino acid has been found only in zein sequences closely related to the cZ22C group (Spena *et al.*, 1982; Geraghty *et al.*, 1982).

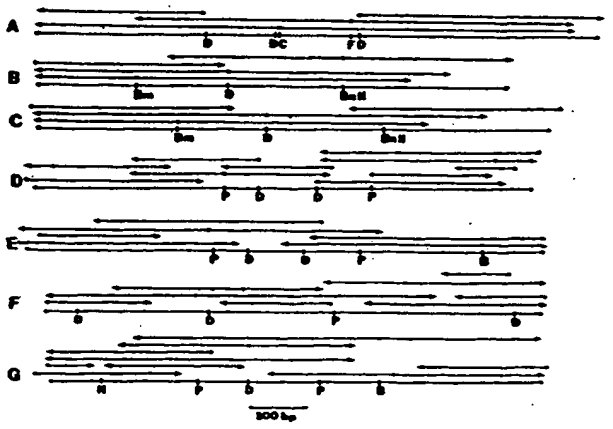


FIG. 1. **Restriction enzyme map and sequencing strategy of zein cDNA clones.** Only selected restriction sites are shown. *Horizontal arrows* indicate the strategy followed for the determination of DNA sequences. The cDNA clones are: *A*, cZ22C2; *B*, cZ19A2; *C*, cZ19B1; *D*, cZ19C1; *E*, cZ19C2; *F*, cZ19D1; and *G*, cZ15A3. Restriction enzymes: *B*, *Bal*I; *Bm*, *Bam*HI; *BnII*, *Ban*II; *C*, *Hinc*II; *D*, *Dde*I; *F*, *Hin*fI; *N*, *Nde*I; and *P*, *Pst*I.

Analysis of M_r 19,000 Zein Sequences—We identified five distinct groups of M_r 19,000 zein clones. Members of the four groups Z19A, Z19B, Z19C, and Z19D did not cross-hybridize at a stringent criterion, whereas those of the fifth group, Z19AB, cross-hybridized to members from Z19A and Z19B. The representative member of the Z19AB group, gZ19AB-1, is a subclone of the genomic clone gZ99, whose sequence was previously reported (Pedersen *et al.*, 1982). The remaining groups are represented by cDNA clones, and their sequences are compared to that of gZ19AB-1 (Fig. 4).

Sequences from the groups Z19A, Z19B, and Z19AB are the most closely related and show 94–95% homology. As was the case with the M_r 22,000 zein clones, the nucleotide differences among these sequences show a fairly random distribution (Fig. 4). This is especially true of the cZ19A-2 and cZ19B-1 sequences; the longest stretch of perfect homology between them is less than 50 nucleotides. However, both of these clones have regions of up to 70 nucleotides that are identical to gZ19AB-1. The differences in the distribution of nucleotide changes between clone groups Z19A, Z19B, and Z19AB may explain their cross-hybridization behavior.

To compare members of the same group we determined the DNA sequences of cZ19C-1 and cZ19C-2 (Fig. 4). These two sequences differ by only one nucleotide in the 5′ noncoding region and by four bases over the first 219 coding nucleotides. From that point to the end of the shorter clone, cZ19C-2, the sequences are identical. The four base differences lead to three conservative amino acid differences (Fig. 5).

To align the sequences of Z19A, Z19B, Z19AB, and Z19C groups, several gaps were introduced (Fig. 4). The largest is a 24-base pair deletion in the Z19A, Z19B, and Z19AB sequences (collectively referred to as Z19A/B) relative to the Z19C sequences. The insertion/deletion events that have taken place during the divergence of the genes corresponding to these clones have rendered the Z19C zeins with six more amino acids than the Z19A/B zeins. The Z19C sequences are on the average 85% homologous to the Z19A/B groups. At the 3′ end the average of 85% homology is maintained. However, in the 5′ noncoding region these sequences are only 55–60% homologous.

To align cZ19D-1 with the other sequences, additional gaps had to be introduced (Fig. 4). The longest is a 24-base pair deletion in Z19D-1. This sequence has diverged more than the others and is only 75% homologous.

Five of the six clones encode complete zein polypeptides (Fig. 5). The remaining clone, cZ19A-2, is missing nucleotides for the first three codons. All of the complete polypeptides contain a 21-amino acid signal peptide and an internal tandemly repeated peptide of approximately 20 amino acids that has been previously described (Argos *et al.*, 1982). The polypeptides range in size from 233 to 240 amino acids. As in the case of the M_r 22,000 zein clones, approximately 45% of the nucleotide differences among these clones result in amino acid substitutions. Most of the resulting changes are conservative; however, at a few positions neutral amino acids have been replaced with charged, and vice versa.

Northern Blot Analysis—Previous analyses revealed that the mRNAs encoding the M_r 22,000 zeins can be resolved from the mRNAs for the M_r 15,000 and 19,000 zeins by gel electrophoresis under denaturing conditions (Wienand and Feix, 1978; Larkins *et al.*, 1983). However, the mRNAs for the

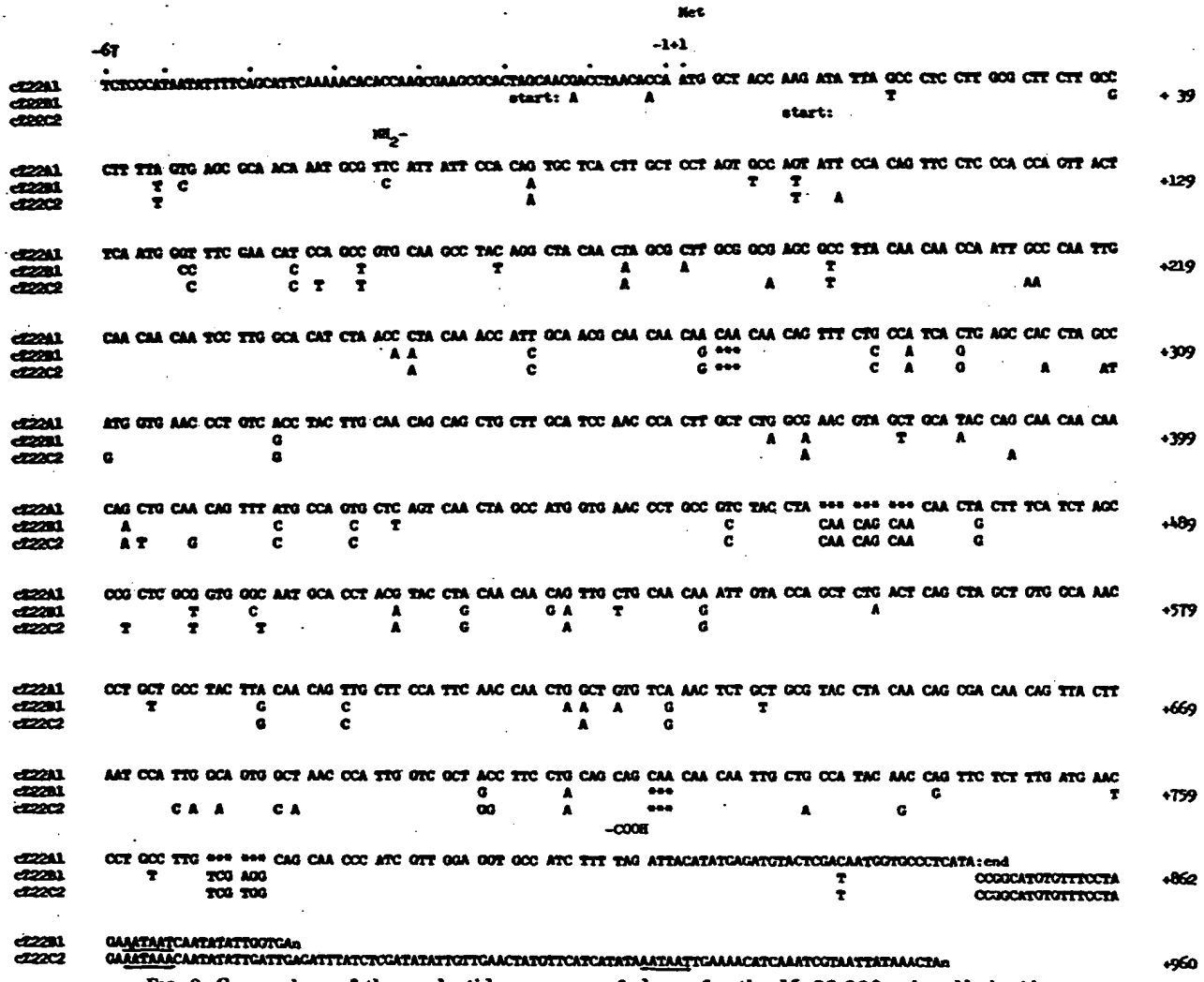


FIG. 2. **Comparison of the nucleotide sequence of clones for the M_r 22,000 zeins.** Nucleotides are numbered starting with the first base of the initiation codon and are indicated on the *right-hand* margin. The sequences of cZ22A1 and cZ22B1 (previously named pZ22.1 and pZ22.3, respectively) were previously reported (Marks and Larkins, 1982). The complete nucleotide sequence of cZ22A1 is given, but only variable nucleotides at corresponding positions are listed for the other clones. *Asterisks* indicate positions where we have introduced gaps in the sequence to maximize homology. The positions of the first nucleotides for cZ22B-1 and cZ22C-2 follow the colons. *An* indicates the position at which a sequence terminates in a poly(A) tail. Positions corresponding to the initiating methionine (Met) and the NH_2 terminus and COOH terminus of the encoded polypeptide are indicated.

FIG. 3. **Comparison of amino acid sequences deduced from DNA sequences of clones for the M_r 22,000 zeins (Fig. 2).** The complete sequence of cZ22A1 is given in the standard single letter amino acid code, but only variable amino acid residues at corresponding positions are listed for the other sequences. The position corresponding to the NH_2 terminus of the mature protein is indicated. Numbering of amino acids begins with the first methionine in the signal peptide. Amino acids are divided into groups of 10 and include gapped positions.

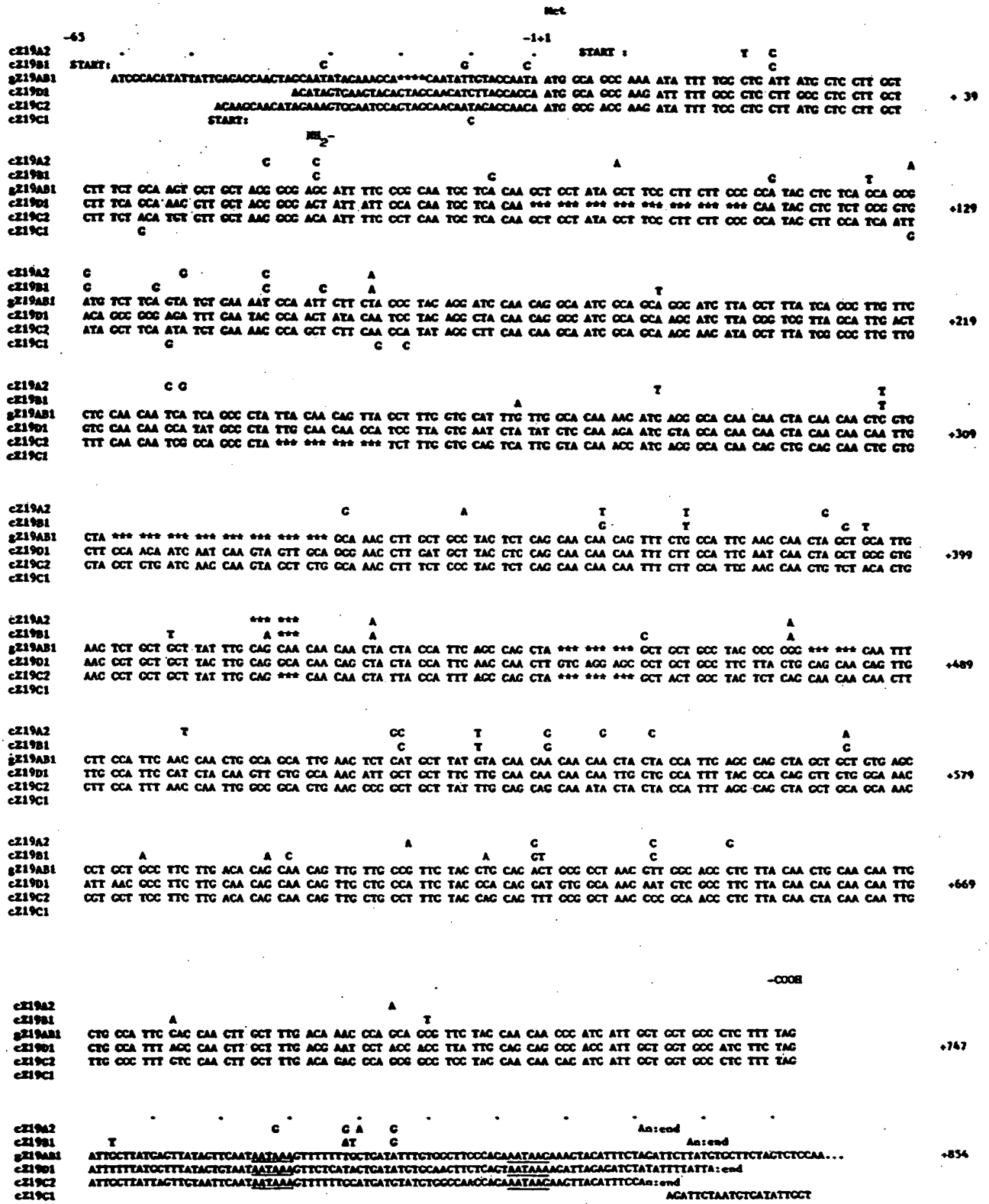


FIG. 4. **Comparison of the nucleotide sequence of clones for the M_r 19,000 zeins.** The nucleotide sequence of gZ19AB1 (previously named ZG99) was previously reported (Pedersen *et al.*, 1982). The complete sequences of gZ19AB1, cZ19D1, and cZ19C2 are given. The sequences of the cZ19A2 and cZ19B1 follow the *colons*, and only nucleotides in the sequences that vary from gZ19AB1 are shown. Likewise, the cZ19C1 sequence begins after the *colon* and only nucleotides in that sequence which vary from cZ19C2 are shown. *An* indicates a sequence terminating in poly(A) tail. *Asterisks* indicate positions where we have introduced gaps in the sequences to maximize homology. The putative polyadenylation signals are *underlined*. Positions corresponding to the initiating methionine (Met) and the NH_2 terminus and COOH terminus of the encoded polypeptide are indicated.

```
cZ19A2  START :   FL          T          T          V          Q                    P
cZ19B1              L         T                     S V        Q
cZ19AB1     MAAKIFCLIN LLGLSASAAT ASIFPQCSQA PIASLLPPYL SPAMSSVCEN PILLPYRIQQ AIAAGILPLS PLFLQQSSAL LQQLPLVHLL AQNIRAQQLQ
cZ19D1      MAAKIFALLA LLALSANVAT ATIIPQCSQ* ********QYL SPVTAARFEY PTIQSYRLQE AIAASILRSL ALTVQQPYAL LQQPSLVNLY LQRIVAQQLQ      100
cZ19C1      MATKIFSLLM LLALSACVAN ATIFPQCSQA PIASLLPPYL PSMIASVCEN PALQPYRLQQ AIAASNIPLS PLLPQQSPAL ****SLVQSL VQTIRAQQLQ
cZ19C2                      T                                 I  I

cZ19A2                    G         H                  * * Q              Q                    A  L      P        D
cZ19B1                                                 GS    S     *  Q              P   Q                    P  L                G    T
cZ19AB1     QLVL******  **ANLAATSQ QQQFLPFNQL AALMSAAYLQ QQQLLPFSQL ***AAAYPR* *QFLPFNQLA ALMSHAIVQQ QQLLPFSQLA AVSPAAFLTQ
cZ19D1      QQLLPTINQV VAANLDAYLQ QQQFLPFNQL AGVNPAAYLQ AQQLLPFNQL VRSPAAFLLQ QQLLPFHLQV VANIAAFLQQ QQLLPFYPQV VGNINAFLQQ      200
cZ19C1      QLVLPLINQV ALANLSPTSQ QQQFLPFNQL STLNPAAYLQ *QQLLPFSQL ***ATAYSQQ QQLLPFNQLA ALNPAAYLQQ QILLPFSQLA AANRASFLTQ
cZ19C2

cZ19A2                  A    A V          N        T
cZ19B1      P      Q V    A               N        V
cZ19AB1     QQLLPFYLHT APNVGTLLQL QQLLPFDQLA LTNPAAFYQQ PIIGGALF
cZ19D1      QQLLPFYPQD VANNVAFLQQ QQLLPFSQLA LTNPTTLLQQ PTIGGAIF                                                              248
cZ19C1      QQLLPFYQQF AANPATLLQL QQLLPFVQLA LTDPAASYQQ HIIGGALF
cZ19C2
```

FIG. 5. **Comparison of amino acid sequences deduced from the DNA sequence of clones for the M_r 19,000 zeins (Fig. 4).** The complete amino acid sequences deduced from gZ19AB1, cZ19D1, and cZ19C1 are given. Only amino acid residues for cZ19A2 and cZ19B1 that differ from gZ19AB1 are shown as are only amino acid residues for cZ19C2 that differ from the sequence of cZ19C1. *Asterisks* indicate gaps in the amino acid sequence. The positions of the NH_2 terminus and COOH terminus of the mature proteins are indicated. The number of amino acids begins with the first methionine in the signal peptide. Amino acids are divided into groups of 10 and include gapped positions.

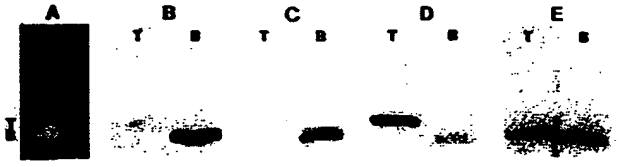


FIG. 6. **Northern blot analysis of zein mRNAs.** *A*, 18 DAP endosperm poly(A) RNA from membrane-bound polyribosomes was resolved on 1.4% agarose gels that contained 5 mM methyl mercury. RNA from gel slices that contained the top (*T*) and bottom (*B*) zein mRNA bands was transferred to nitrocellulose and hybridized to nick-translated zein cDNA inserts as described under "Materials and Methods." The cDNA probes were: *B*, cZ15A3; *C*, cZ19C1; *D*, cZ22B1; and *E*, gZ19AB1.

M_r 15,000 and 19,000 zeins have not been resolved. We used a modified Northern blot analysis to determine if our clone groups correspond to distinct mRNA populations encoding the M_r 22,000, 15,000, or 19,000 zeins.

Poly(A) rRNA was isolated from membrane-bound polysomes of maize endosperm and resolved on a methyl mercury agarose gel (Fig. 6*A*). The upper and lower bands corresponding to zein mRNAs were excised, blotted onto separate nitrocellulose filters, and hybridized to representative clones. Probes representing the Z15 and Z19C groups hybridized only to the lower band (Fig. 6, *B* and *C*). The cZ22B-1 probe hybridized only to the upper band under conditions that allowed cross-hybridization between the three closely related M_r 22,000 zein groups (Fig. 6*D*). These results indicate that the mRNA populations corresponding to the respective clone groups are homogenous and probably encode only a single size class of protein. The 19AB probe hybridized to both the upper and lower bands (Fig. 6*E*), a result which is consistent with the existence of two subsets of mRNAs corresponding to this group (Hu *et al.*, 1982; Heidecker and Messing, 1983). Presumably, one group of mRNAs encodes M_r 19,000 zeins and the other encodes M_r 22,000 zeins.

Analysis of a M_r 15,000 Zein cDNA Clone—We recently compared the sequence of a cDNA clone and a genomic clone encoding a M_r 15,000 zein.[1] The cDNA clone lacks a portion of the 5′ coding region but contains the entire 3′ noncoding sequence of its template mRNA. Except for one nucleotide difference, the cDNA sequence is identical to the genomic sequence for its entire length.

We report here the sequence of an additional cDNA clone that encodes a complete M_r 15,000 zein polypeptide (Fig. 7). This sequence differs from the genomic clone[1] by a few nucleotides. Since we estimated that there may only be two genes for the M_r 15,000 zeins (Wilson and Larkins, 1984), these two sequences may represent them. However, some of the base differences between the clones may be due to sequencing artifacts. Because of the high G + C content, many regions of the sequencing gels were compressed, making the sequence determination extremely difficult. Essentially all of the base differences were in these regions.

Previous S1-mapping data suggested that the 5′ end of the mRNA for the M_r 15,000 gene began 65–70 nucleotides before the start codon (see Fig. 7). However, this larger cDNA clone has over 150 noncoding nucleotides at the 5′ end that align perfectly with the 5′ noncoding sequence of the genomic clone. The existence of this cDNA clone indicates that transcription may originate from two regions of the gene. To investigate this, we performed an S1-mapping analysis. A *Hin*dIII-*Dde*I fragment that contains the 5′ end of the mRNA sequence was hybridized to poly(A) RNA isolated from staged endosperms (Fig. 8*A*) and the hybrids digested with S1 nuclease. The majority of protected fragments ended in the same region previously shown to be the 5′ end of the mRNA. However, upon longer exposure of the gel we detected some transcripts that protected the complete noncoding region of the longer cDNA clone (Fig. 8*B*). While these results are consistent with the possibility of a major and minor origin of transcription, they do not exclude processing of a larger transcript for this gene. The longer transcript appears to be more abundant at early stages of endosperm development, as is indicated by the greater protection of the larger fragment with poly(A) RNA from 12 DAP endosperms than from later stages (Fig. 8*B*). Further analysis of the transcriptional activity of the M_r 15,000 gene will be presented elsewhere.[3]

The last 23 nucleotides of the 3′ end of the cZ15A3 were not homologous to the genomic clone. The terminal 28 nucleotides were found to be an inverted copy of the sequence

[3] R. S. Boston and B. A. Larkins, manuscript in preparation.

```
cZ15A3  AATCCATGTACTATGTTTCCACGTCATGCAACGCAACATTCCAAAACCATGGATCATCTATAAATGGCTAGCTCCCACATATGAACTAGTCTCTATCATCATCCAATCCAGATCAGCAAA  - 39
                                                                                                  XXX

cZ15A3  GCGGCAGTGCGTAGAGAGGATCGTCGAACCAGAACAGC ATG AAG ATG GTC ATC GTT CTC GTC GTG TGG CTG GCT CTG TCA GCT GCC AGC GCC TCT GCA  + 60
                                            -1 +1
                                               M  K  M  V  I  V  L  V  V  W  L  A  L  S  A  A  S  A  S  A
        NH2-

cZ15A3  ATG CAG ATG CCC TGC CCC TGC GCG GGG CTG CAG GGC TTG TAC GGC GCT GGC GCC GGC CTG ACG ACG ATG ATG GGC GCC GGC GGG CTG TAC  +150
         M  Q  M  P  C  P  C  A  G  L  Q  G  L  Y  G  A  G  A  G  L  T  T  M  M  G  A  G  G  L  Y

cZ15A3  CCC TAC GCG GAG TAC CTG AGG CAG CCG CAG TGC AGC CCG CTG GCG GCG GCG CCC TAC TAC GCC GGG TGT GGG CAG ACG AGC GCC ATG TAC  +240
         P  Y  A  E  Y  L  R  Q  P  Q  C  S  P  L  A  A  A  P  Y  Y  A  G  C  G  Q  T  S  A  M  Y

cZ15A3  CAG CCG CTC CGG CAA CAG TGC TGC CAG CAG CAG ATG AGG ATG ATG GAC GTG CAG TCC GTC GCG CAG CAG CTG CAG ATG ATG ATG CAG CTT  +330
         Q  P  L  R  Q  Q  C  C  Q  Q  Q  M  R  M  M  D  V  Q  S  V  A  Q  Q  L  Q  M  M  M  Q  L

cZ15A3  GAG CGT GCC GCT ACC GCC AGC AGC AGC CTC TAC GAG CCA GCT CTG ATG CAG CAG CAG CAG CAG CTG CTG GCA GCC CAA GGT CTC AAC CCC  +420
         E  R  A  A  T  A  S  S  S  L  Y  E  P  A  L  M  Q  Q  Q  Q  Q  L  L  A  A  Q  G  L  N  P

cZ15A3  ATG GCC ATG ATG ATG GCG CAG AAC ATG CCG GCC ATG GGT GGA CTC TAC CAG TAC CAG TAC CAG CTG CCC AGC TAC CGC ACC AAC CCC TGT  +510
         M  A  M  M  M  A  Q  N  M  P  A  M  G  G  L  Y  Q  Y  Q  Y  Q  L  P  S  Y  R  T  N  P  C

                                        -COOH

cZ15A3  GCG GTC TCC GCT GCC ATT CCG CCC TAC TAC TGA TTCATGATATTTGGGAAATCTCCTCTATCCATCCCTCTCTATCTATATATGTAATAATGCAGTAAGACCGACACA  +617
         C  V  S  A  A  I  P  P  Y  Y  *

cZ15A3  CATTATCATGTGTGGTATGACCAATAATATATGCATCATAATAAAGTTTTGGTCATACCACACATGATAATGTGTG
```

FIG. 7. **Nucleotide sequence of a cDNA clone encoding a M_r 15,000 zein protein.** The predicted amino acid sequence is shown below the nucleotide sequence. *XXX* indicates the major transcriptional start site, and the *underscored* sequence corresponds to the putative polyadenylation signal. The *overscored* region indicates the position of an inverted nontandem repeated sequence. The NH_2-terminal and COOH-terminal positions of the mature polypeptide are indicated.

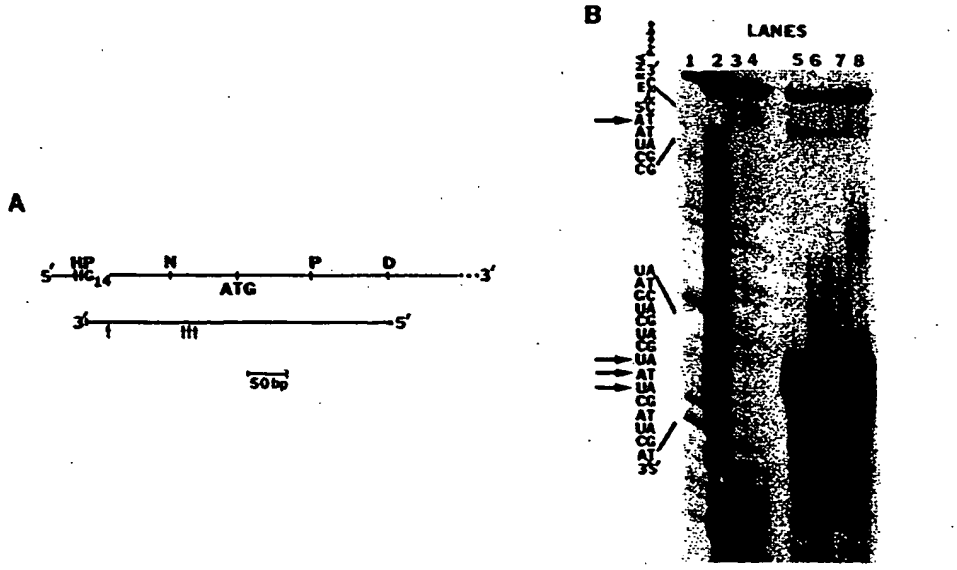


FIG. 8. **S1 mapping of the 5′ region of cZ15A3.** *A*, a partial restriction enzyme map of the 5′ end of cZ15A3 and a region of the clone used in the S1-mapping experiment. G_{14} indicates a G tract of 14 bases that was added during the cloning procedure. The restriction enzyme sites upstream from the G tract are in the poly linker region of the plasmid vector pUC8. The probe was generated by end labeling a *Dde*I fragment with subsequent cleavage at the *Hin*dIII site in the poly linker. The *vertical arrows* indicate the limits of S1 digestion as shown in *B*. *B*, Maxam and Gilbert sequencing reactions of the probe are shown in *lane 1* (G reaction), *lane 2* (G + A reaction), *lane 3* (C + T reaction), and *lane 4* (C reaction). The probe was hybridized to 1 μg of poly(A) RNA from 12 DAP (*lane 5*), 18 DAP (*lane 6*), 22 DAP (*lane 7*), and 28 DAP (*lane 8*) endosperms as described under "Materials and Methods." After the hybridization the probe was digested with S1 nuclease and resolved on a sequencing gel. The sequence in the region of the nuclease-protected bands and the location of the bands (*arrows*) are given on the *left-hand* margin. Restriction enzyme key: *D*, *Dde*I; *H*, *Hin*dIII; *N*, *Nde* I; and *P*, *Pst*I.

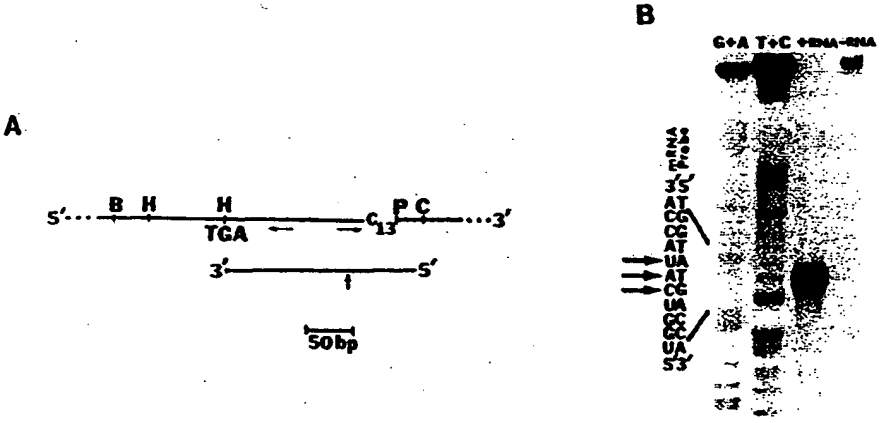


FIG. 9. **Mapping of the 3′ region of cZ15A3.** *A*, a partial restriction enzyme map of the 3′-terminal of cZ15A3 and the region of the clone used in the S1-mapping experiment. The C_{13} represents a C tract of 13 bases that was added during the cloning procedure (Marks *et al.*, 1985). The restriction enzyme sites downstream from the C tract are in the poly linker region of the plasmid vector, pUC8. The probe was generated by cleavage with *Hin*fI, and it was 3′ end labeled by using Klenow fragment. The labeled fragment was recut with *Hin*cII. The *horizontal arrows* show the position of inverted repeats (see Fig. 6), and the *vertical arrows* indicate the limit of S1 digestion as shown in *B*. *B*, Maxam and Gilbert sequencing reactions of the probe are shown in the *first two lanes* as indicated. The probe was hybridized to 1 μg of 18 DAP endosperm poly(A) RNA plus 200 μg tRNA (+RNA) or 200 μg of tRNA without poly(A) RNA (−RNA) as described under "Materials and Methods." After hybridization the probe was digested with S1 nuclease and resolved on a sequencing gel. The sequence in the region of the nuclease-protected bands and the location of the bands (*arrows*) are given at the *left-hand* margin. Restriction enzymes: *B*, *Bal*I; *C*, *Hin*cII; *H*, *Hin*fI; and *P*, *Pst*I.

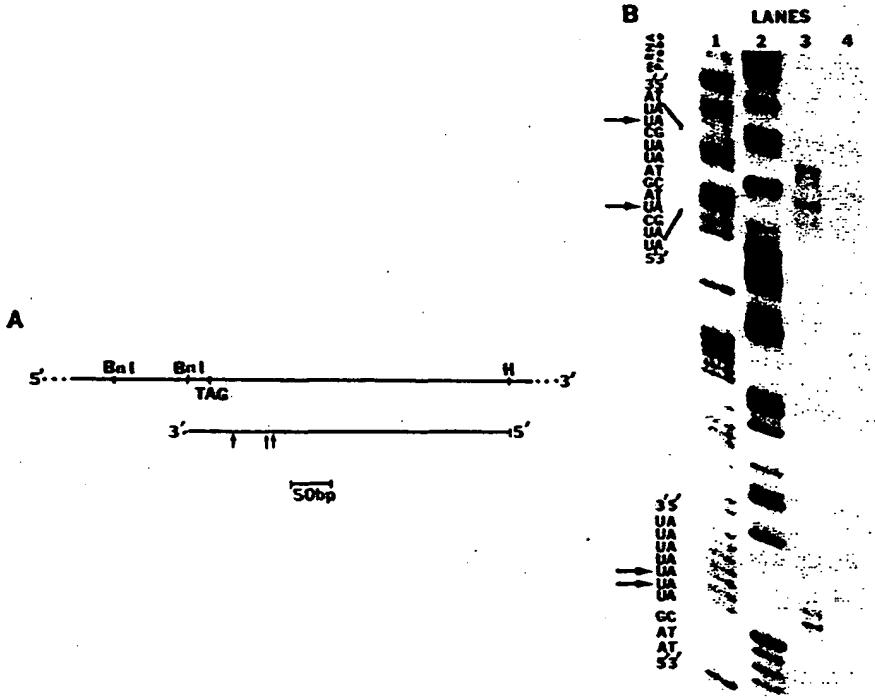


FIG. 10. **Determination of the 3′ terminus of mRNAs closely related to the Z19A/B clones.** *A*, a partial restriction enzyme map of the 3′ coding and noncoding sequence of the clone gZ19AB-1 (also called ZG99, Pedersen *et al.*, 1982) and the region of the clone used in a S1-mapping experiment. The probe was made by cleavage with *Ban*I followed by 3′ end labeling with Klenow fragment and a second restriction digestion with *Hin*cII. The *vertical arrows* indicate the limits of S1 digestion as shown in *B*. *B*, Maxam and Gilbert sequencing reactions of the probe are shown in *lane 1* (G + A reaction) and *lane 2* (C + T reaction). The probe was hybridized to 1 μg of 18 DAP endosperm poly(A) RNA plus 200 μg of tRNA (*lane 3*) and to 200 μg of tRNA alone (*lane 4*) as described under "Materials and Methods." After hybridization the probe was digested with S1 nuclease and resolved on a squencing gel. The sequences in the region of the nuclease-protected bands and the location of the bands (*arrows*) are given on the *left-hand* margin of *B*. Restriction enzymes: *Bal*I, *Ban*I and *H*, *Hin*cII.

found between nucleotides 614 and 642 (Fig. 7). To determine if the 3′ end of the cDNA is authentic, we used a *Hin*fI to *Hin*cII fragment 3′ end labeled at the *Hin*fI site to probe endosperm poly(A) RNA (Fig. 9*A*). The only protected fragments extended a few bases into the terminal inverted repeat (Fig. 9*B*). Thus, this inverted repeat is apparently a cloning artifact and is not due to sequence heterogeneity.

Analysis of Polyadenylation—Many of the M_r 19,000 zein genes were found to contain two polyadenylation signals (Messing *et al.*, 1983). To quantitate the relative utilization of these signals in the genes belonging to the Z19A/B groups (Fig. 4) we did an S1-mapping experiment. A *Ban*I to *Hin*cII fragment from the genomic clone 19AB-1 was 3′ end labeled on the *Ban*I site (Fig. 10*A*) and hybridized to endosperm poly(A) RNA (Fig. 10*B*). The probe should be protected by mRNA corresponding to the Z19A/B groups since these sequences differ by only a few nucleotides in this region. Two major fragments that differ in size by seven nucleotides were protected (Fig. 10*B*). Both of these fragments extend past the second polyadenylation signal. Furthermore, the shorter and

longer fragments end precisely at the last nucleotides found in cZ19A-2 and cZ19B-1, respectively (Fig. 4). Several additional bands, fainter in intensity, correspond to fragments ending three bases after the first polyadenylation signal. Since addition of the poly(A) tail normally occurs within 10–30 nucleotides of the polyadenylation sequence (Nevins, 1983), these may not represent authentic 3′ ends of the transcripts. They may instead represent a region of instability due to a long tract of Ts in this area. A few faint bands corresponding to mRNAs terminating 10–40 nucleotides beyond the first polyadenylation signal also appeared.

Heidecker and Messing (1983) have sequenced two cDNAs closely related to 19A/B groups that terminate after the first polyadenylation signal. Thus, the first site is probably used, but at a much lower frequency than the second. Similar results were obtained with probes from the 19C and 19D groups (data not shown).

DISCUSSION

To study the heterogeneity among the zein mRNAs in the inbred W64A, we characterized the structure of a number of zein cDNA clones. Each of the three groups of M_r 22,000 zein sequences is 60–70% homologous to the M_r 19,000 zein sequences. Sequences among the M_r 22,000 groups showed 92% homology, and sequences among the M_r 19,000 groups have homology ranging from 75 to 95%. The relationships among all these zein sequences are illustrated in Fig. 11, which suggests the order of duplications that gave rise to these groups.

Hu *et al.* (1982) characterized a genomic clone, Z4, closely related to our Z19A/B groups that contains a large internal duplication. The duplication adds 30 amino acids to the encoded polypeptide and makes it the size of a M_r 22,000 zein. By Northern blot analysis we found that in the W64A inbred this type of duplication appears to be restricted to genes within the Z19A/B group (Fig. 6).

The duplication event suggests that the M_r 22,000 zeins may have originated from a M_r 19,000 gene, and close examination of the M_r 22,000 sequences provides evidence for this. The sequence from nucleotides 264–396 is 85% homologous to the sequence from nucleotides 397–540. The duplication that formed this region would have added approximately 40 amino acids to the encoded polypeptide. A series of base substitutions would account for the subsequent diversion of the genes into two distinct gene families.

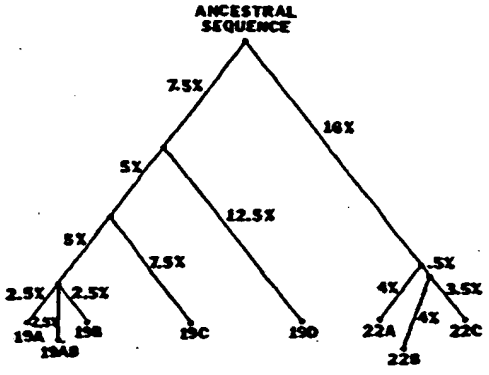


FIG. 11. Comparison of sequence homology among zein mRNAs. The phylogeny was reconstructed with the method described by Fitch and Margoliash (1967) from the approximate per cent of nucleotide substitutions that have taken place among the zein genes. The percentages listed are the approximate per cent of nucleotide substitution between putative gene branching points.

We have also sequenced a cDNA clone for the M_r 15,000 zein which is nearly identical to that obtained for a genomic clone encoding this protein.[1] The sequences of the M_r 15,000 zeins show no homology to those for either the M_r 19,000 or 22,000 zeins, indicating that they comprise a separate family of genes. The M_r 15,000 zeins have a much higher G + C content, lack an internal repeating nucleotide sequence, and are present in only one or two copies in the genome (Wilson and Larkins, 1984). According to S1 mapping, the mRNA encoding the M_r 15,000 zein primarily begins 65–70 nucleotides before the start codon (Fig. 8). The sequence of cZ15A3 differs from this by having a 5′ noncoding sequence of over 150 nucleotides. Based on S1 mapping we estimate that the longer mRNA is 100–1000-fold less abundant than the major mRNA species (Fig. 8). Typical "TATA" boxes (Efstratiadis *et al.*, 1980) are found in the genomic clone just upstream from the start of both mRNA species suggesting that both start sites may be used. But it is also possible that the larger transcript is a precursor of the smaller one. The existence of zein genes with two transcriptional promoters is not without precedence. Langride and Feix (1983) analyzed a gene that encodes a M_r 22,000 zein that appears to have two promoters. However, these promoters are separated by over 1000 nucleotides, and they have nearly equal levels of transcriptional activity.

Of the five zein cDNA clones that contained poly(A) tails only three had the consensus sequence "AATAAA" which is found on the 3′ terminus of many animal genes (Nevins, 1983). This sequence was located 50–100 nucleotides before the start of the poly(A) tail, whereas in animal genes the polyadenylation signal is generally 10–30 nucleotides upstream from the poly(A) tail. Messing *et al.* (1983) identified several putative variants of the polyadenylation signal in zein cDNA sequences that were located closer to the poly(A) tail. In the A20 and A30 clones, which are closely related to our Z19C and C19A/B groups, respectively, they found the sequence "AATAAG." We also found this sequence on the 3′ terminus of the Z19C and Z19A/B clones (Fig. 4). In their B49 sequence (closely related to the Z22C group) another putative variant, "AATAAT," was found (Geraghty *et al.*, 1982). This sequence is also found on the 3′ terminus of our Z22B and Z22C clones (Fig. 2).

In addition to the variation in the sequence of the polyadenylation signal, some plant genes also differ from animal genes by having two potential polyadenylation sites (Messing *et al.*, 1983). Hu *et al.* (1982) reported the sequence of a genomic clone, Z4, that has a "AATAAA" sequence and a "AATAAG" sequence 26 and 65 bases, respectively, downstream from the stop codon. They also isolated cDNA clones that terminated in a poly(A) tail after each of these sequences (Heidecker and Messing, 1983). These clones are closely related to our cZ19A2 and cZ19B1 clones which end in a poly(A) tail after the second putative polyadenylation signal.

Our results indicated that most of the zein mRNA that corresponds to the Z19A/B clones terminated after the second polyadenylation sequence. This appears to be in contrast to mRNAs closely related to the M_r 22,000 clone groups (Z22A, B, C). Three of four of the cDNA clones which belong to these groups represented mRNAs that were polyadenylated after the first polyadenylation sequence (Geraghty *et al.*, 1982; Spena *et al.*, 1982). The remaining clone, cZ22C-2, represented an mRNA that was polyadenylated after a second polyadenylation site (Fig. 2). Interestingly, the sequence that Spena *et al.* (1982) determined for the cDNA clone, za1, is identical to that of cZ22C2 up to the point at which za1 ends in a poly(A)

tail. Thus, it appears as though there are preferred polyadenylation sites, but their utilization is variable.

REFERENCES

Argos, P., Pedersen, K., Marks, M. D., and Larkins, B. A. (1982) *J. Biol. Chem.* **257**, 9984–9990

Bailey, J. M., and Davidson, N. (1976) *Anal. Biochem.* **70**, 75–85

Berk, A. J., and Sharp, P. A. (1977) *Cell* **12**, 721–732

Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slighton, J. L., Blechl, A. E., Smithies, O., Barrelle, E. E., Shoulders, C. C., and Proudfoot, N. J. (1980) *Cell* **21**, 653–668

Fitch, W. M., and Margoliash, E. (1967) *Science* **155**, 279–284

Geraghty, D. E., Messing, J., and Rubenstein, I. (1982) *EMBO J.* **1**, 1329–1335

Heidecker, G., and Messing, J. (1983) *Nucleic Acids Res.* **11**, 4891–4906

Hu, N.-T., Peifer, M. A., Heidecker, G., Messing, J., and Rubenstein, I. (1982) *EMBO J.* **1**, 1337–1342

Krayev, A. S., Kramerov, D. S., Skryabin, K. G., Ryskov, A. P., Bayer, A. A., and Georgiev, G. P. (1980) *Nucleic Acids Res.* **8**, 1201–1215

Langride, P., and Feix, G. (1983) *Cell* **34**, 1015–1022

Larkins, B. A., Pedersen, K., Marks, M. D., Wilson, D. R., and Argos, P. (1983) in *Structure and Function of Plant Genomes* (Ciferri, O., and Dure, L., III, eds) pp. 78–83, Plenum Publishing Corp., New York

Maniatis, T., Hardison, R. G., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1976) *Cell* **15**, 687–701

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp. 207–209, Cold Spring Harbor Laboratory, New York

Marks, M. D., and Larkins, B. A. (1982) *J. Biol. Chem.* **257**, 9976–9983

Marks, M. D., Lindell, J. S., and Larkins, B. A. (1985) *J. Biol. Chem.* **260**, 16445–16450

Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560

Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J., and Rubestein, I. (1983) in *Genetic Engineering of Plants* (Hollaender, A., Kosuge, T., and Meredith, C. P., eds) pp. 211–228, Plenum Publishing Corp., New York

Nevins, J. R. (1983) *Annu. Rev. Biochem.* **52**, 441–466

Pedersen, K., Devereux, J., Wilson, D. R., Sheldon, E., and Larkins, B. A. (1982) *Cell* **29**, 1015–1026

Spena, A., Viotti, A., and Pirrotta, V. (1982) *EMBO J.* **1**, 1589–1594

Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 5201–5205

Weinand, U., and Feix, G. (1978) *Eur. J. Biochem.* **92**, 605–611

Wilson, D. R., and Larkins, B. A. (1984) *J. Mol. Evol.* **20**, 330–340

EXHIBIT C

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 7094–7097, June 1997
Plant Biology

Expression of a mutant α-zein creates the *floury2* phenotype in transgenic maize

(endosperm/prolamin)

CRAIG E. COLEMAN*†, AMY M. CLORE*, JERRY P. RANCH‡, REGINA HIGGINS‡, MAURICIO A. LOPES§, AND BRIAN A. LARKINS*¶

*Department of Plant Sciences, University of Arizona, Tucson, AZ 85721; ‡Pioneer Hi-Bred International, Inc., 7300 NW 62nd Avenue, Johnston, IA 50131-1004; and §Centro Nacional de Pesquisa de Milhoe Sorgo–Empresa Brasileira de Pesquisa Agropecuaria, Caixa Postal 151, 35700 Sete Lagoas, MG, Brazil

Contributed by Brian A. Larkins, May 6, 1997

ABSTRACT **The maize *floury2* mutation results in the formation of a soft, starchy endosperm with a reduced amount of prolamin (zein) proteins and twice the lysine content of the wild type. The mutation is semidominant and is associated with small, irregularly shaped protein bodies, elevated levels of a 70-kDa chaperone in the endoplasmic reticulum, and a novel 24-kDa polypeptide in the zein fraction. The 24-kDa polypeptide is a precursor of a 22-kDa α-zein protein that is not properly processed. The defect is due to an alanine-to-valine substitution at the C-terminal position of the signal peptide, which causes the protein to be anchored to the endoplasmic reticulum. We postulated that the phenotype associated with the *floury2* mutation is caused by the accumulation of the 24-kDa α-zein protein. To test this hypothesis, we created transgenic maize plants that produce the mutant protein. We found that endosperm in seeds of these plants manifests the *floury2* phenotype, thereby confirming that the mutant α-zein is the molecular basis of this mutation.**

Zeins are prolamin storage proteins that accumulate in the endosperm of maize (*Zea mays* L.) seeds. They are composed of four different types of polypeptides, classified as α-, β-, γ-, and δ-zeins (1). Accretions of zein proteins form spherical protein bodies within the lumen of the endoplasmic reticulum (ER), and there is a distinct spatial arrangement of these proteins within a protein body: β- and γ-zeins are located on the periphery, whereas α- and δ-zeins are found in the interior (2, 3). Collectively, the zein proteins are rich in glutamine and proline, but they lack lysine and tryptophan. Because zeins constitute such a large proportion of the total seed protein (60–70%), the amino acid composition of these proteins causes the grain to be of inferior nutritional quality for monogastric animals.

Efforts to improve the protein quality of maize seed have focused on mutants in which zein synthesis is reduced and the lysine content is increased. The first "high-lysine" mutants to be identified were opaque2 (*o2*) and floury2 (*fl2*) (4, 5). Unfortunately, the favorable nutritional quality of these mutants is offset by the inferior physical properties of their soft, starchy endosperms. It appears that the starchy endosperm of the *o2* and *fl2* mutants is caused by changes in the nature of their protein bodies. The *o2* mutation affects a transcriptional activator of a subset of α-zein genes, leading to a reduction in α-zein protein synthesis and the formation of protein bodies that are significantly smaller than normal (6–9). The *fl2* mutation, which is semidominant, causes a decrease in synthesis of all classes of zeins, and the resultant protein bodies are not only smaller than normal, but they are also asymmetrical and misshapen (10, 11). Another feature of *fl2* endosperm is the overexpression of the ER-resident binding protein (BiP), which becomes deposited at the periphery of the mutant protein bodies (12–15).

We have postulated that the phenotype associated with *fl2* is caused by the accumulation of a novel 24-kDa α-zein protein (16, 17). This hypothesis is partially based on tight linkage between the gene encoding the 24-kDa protein and the *fl2* locus, but it is also consistent with the abnormal structure of the protein. The mutant protein is 2 kDa larger than expected, because of a defect in its processing following targeting to the ER lumen. An alanine-to-valine substitution at the C-terminal position of the signal peptide prevents its removal, thereby anchoring the protein to the lumenal face of the ER membrane (18). To investigate whether this protein is responsible for the mutant phenotype, we transformed normal maize plants with the gene encoding the mutant 24-kDa α-zein protein. We show here that seeds of these plants manifest the key phenotypic characteristics associated with the *fl2* mutation.

MATERIALS AND METHODS

Transformation of Maize Plants. Maize embryos were cotransformed with the 24-kDa α-zein gene in plasmid pCC515 (17) and the bacterial bialaphos (*BAR*) gene (19) as a selectable marker. The 24-kDa α-zein gene in pCC515 is flanked by 3.0 kb of 5′ and 3.7 kb of 3′ noncoding sequences. The selectable marker gene plasmid is of the form *ubi::ubiintron::BAR::pinII*, where *ubi* is a maize ubiquitin promoter and first intron and *pinII* is the potato protease inhibitor II 3′ noncoding sequence. Plasmid DNAs were delivered by microprojectile bombardment to embryogenically responsive, immature embryos from the maize variety High Type II (20). Embryos were recovered from herbicide-resistant calli and grown to maturity. Transgenic plants were outcrossed as the female parent to an inbred line, and the progeny were scored for the floury kernel phenotype.

PCR Amplification of the 24-kDa α-Zein Gene. Genomic DNA was extracted from seedlings germinated from normal and floury kernels (21). The DNA was amplified by PCR using primers corresponding to the signal peptide coding sequences of the 24-kDa α-zein gene (5′-GCCCTTTTAGTGAGCGCAACAAATGTG-3′) and coding sequences for the seventh α-helical repeat of the protein (5′-GCAGGGTTTGCCATAGCTAGCTGATG-3′). Products were separated by electrophoresis in 1% agarose gels, stained with ethidium bromide, and photographed with a Polaroid DS-34 camera.

Protein Extraction from Maize Flour and Immunoblotting. A portion of each endosperm was cut from the seed prior to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: BiP, binding protein; ER, endoplasmic reticulum.

†Present address: Department of Botany, Brigham Young University, Provo, UT 84602.
¶To whom reprint requests should be addressed.

germination and converted into a fine flour using a ball mill. Proteins were extracted from the meal and separated according to their solubility in 70% alcohol (22). Alcohol-soluble (zein) proteins were separated by SDS/PAGE, blotted onto nitrocellulose, and immunoreacted with rabbit anti-α-zein polyclonal antibody (23). Alcohol-insoluble proteins were separated by SDS/PAGE, blotted, and immunoreacted with a rabbit anti-BIP polyclonal antibody (12). Goat anti-rabbit alkaline phosphatase conjugate was used for indirect detection of α-zein and BiP on the immunoblots (24).

Fixation and Embedding of Endosperms and Electron Microscopy. Seeds were harvested 18 days after pollination from a self-fertilized maize plant that was hemizygous for the 24-kDa α-zein transgene. Endosperms were fixed, embedded, sectioned, and viewed with a transmission electron microscope as described elsewhere (11).

RESULTS

Detection of the Transgene in Transformed Maize Seedlings. Transgenic maize plants were generated by a biolistic method using microprojectiles coated with plasmid pCC515 (17), which contains the 24-kDa α-zein gene within a genomic DNA fragment, and a plasmid containing the *BAR* gene (20), which confers resistance to the herbicide Basta. Plants from 25 herbicide-resistant events that were recovered from the transformation were crossed as females to an untransformed inbred line. F_1 progeny from 17 of these crosses segregated approximately 1 to 1 for floury-appearing kernels, consistent with the presence of a single site of transgene integration. To determine whether this phenotype was associated with the insertion of the 24-kDa α-zein gene, we used PCR primers for the coding region of the 24-kDa α-zein gene to amplify genomic DNA from F_1 seedlings. A 560-bp fragment was produced from DNA of the floury seedlings (Fig. 1*A*, lanes 4, 6, and 8), but not the wild-type seedlings (Fig. 1*A*, lanes 3, 5, and 7). A fragment of similar size was amplified from W64A*fl2* DNA (Fig. 1*A*, lane 9), but not from DNA of untransformed seedlings (Fig. 1*A*, lanes 1 and 2).

Immunodetection of α-Zein Proteins from Seeds of Transgenic Plants. An immunoblot of α-zein proteins from mature seeds was prepared to determine whether insertion of the 24-kDa α-zein gene resulted in synthesis of this protein (Fig. 1*B*). The blot shows that the appearance of a 24-kDa protein band, indicated by arrowheads in lanes 4, 6, and 8 of Fig. 1*B*, was always associated with the floury phenotype and insertion of the transgene. The 24-kDa protein band in these samples is similar to a band of identical molecular mass that was found in W64A*fl2* zein (indicated by an arrowhead in lane 9 of Fig. 1*B*). The 24-kDa α-zein was not present in samples from normal progeny or seeds of untransformed plants (Fig. 1*B*, lanes 1, 2, 3, 5, and 7). Three bands with molecular masses greater than 24 kDa were detected in all of the samples. In W64A*fl2*, these polypeptides are glycosylated forms of 19-kDa α-zein proteins that are products of a gene(s) closely linked to the *fl2* locus (J. W. Gillikin and R. S. Boston, personal communication). Apparently, the same glycoproteins are present in the maize line used for the transformation, although they had no effect on the kernel phenotype.

Analysis of BiP Expression in Seeds of Transgenic Plants. One phenotypic characteristic associated with the *fl2* mutation is overexpression of the 70-kDa ER-resident chaperone, BiP. Using anti-BiP antibody, an immunoblot of the alcohol-insoluble proteins showed that the amount of BiP in mature

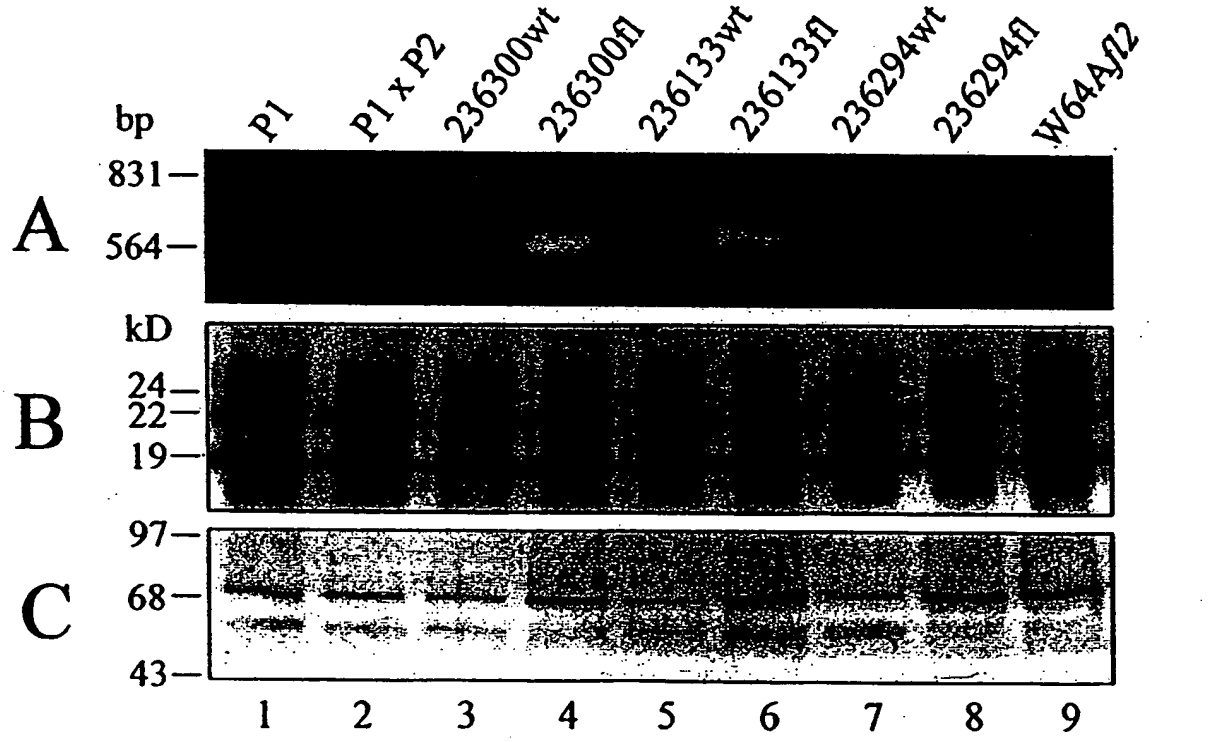


FIG. 1. Transformation of the 24-kDa α-zein gene into transgenic maize plants leads to synthesis of the encoded protein and overexpression of BiP. Analysis of 3 of 17 independent transgenic lines is shown. One normal and one floury kernel from transgenic lines 236300 (lanes 3 and 4), 236133 (lanes 5 and 6), and 236294 (lanes 7 and 8) were analyzed and compared with kernels from the untransformed (lane 1, P1), the untransformed parent outcrossed to a normal inbred (lane 2, P1 × P2), and W64A*fl2* (lane 9) plants. (*A*) Using PCR primers specific to the 24-kDa α-zein gene, a 560-bp DNA fragment was amplified from genomic DNA of seedlings germinated from floury kernels, but not wild-type kernels. (*B*) An immunoblot of α-zeins shows a 24-kDa protein band (indicated by arrowheads) in samples from floury kernels, but not wild-type kernels. (*C*) An immunoblot of the 70-kDa maize homolog of BiP.

transgenic kernels was comparable to that of W64A*fl2* (Fig. 1*C*, lanes 4, 6, 8, and 9). Normal quantities of BiP were detected in wild-type siblings and untransformed kernels (Fig. 1*C*, lanes 1, 2, 3, 5 and 7).

Protein Body Structure in the Endosperms of Seeds of Transgenic Plants. Another phenotypic characteristic associated with the *fl2* mutation is the formation of misshapen protein bodies. Protein bodies in the developing endosperm of seeds not expressing the 24-kDa α-zein gene are circular in cross-section and are relatively discreet; similar to the wild type (Fig. 2*A*; ref. 2), whereas those from seeds expressing the gene have a convoluted shape and aggregate into large masses of protein, similar to W64A*fl2* (Fig. 2*B*; ref. 11). Analysis at higher magnification revealed that protein bodies in the developing endosperm of seeds not expressing the 24-kDa α-zein gene contain darkly staining zein proteins primarily on the periphery of the protein bodies, as is seen in wild type (Fig. 2*C*, arrowheads; ref. 2). In contrast, those from endosperms expressing the gene have many regions of darkly staining zein in their interior (Fig. 2*D*, arrowheads; ref. 11), indicating an abnormal organization of α-, β-, and γ-zeins within the protein body.

DISCUSSION

Previous studies provided strong genetic evidence implicating a mutant 24-kDa α-zein protein as the cause of the phenotype

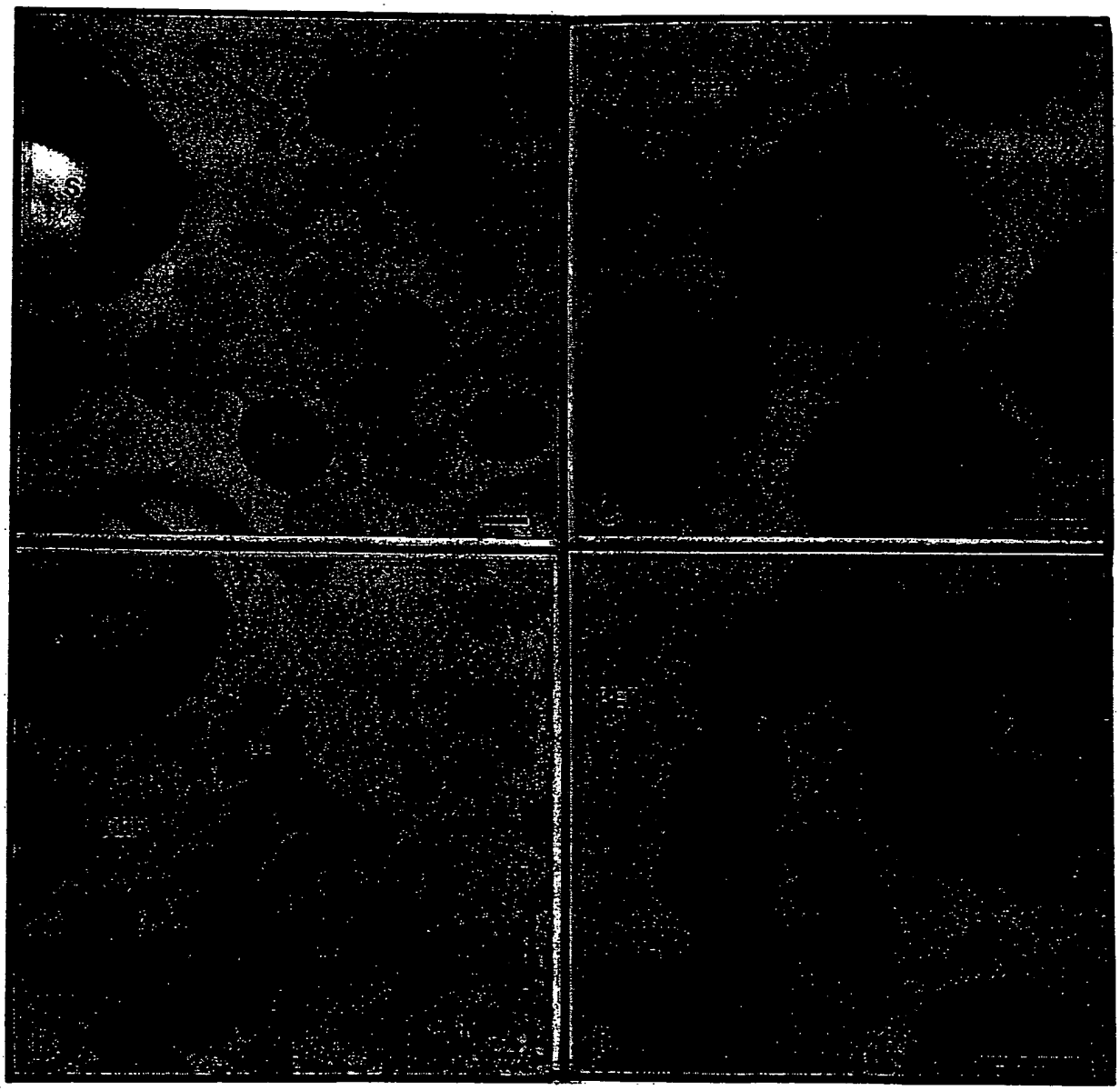


FIG. 2. Comparison of protein bodies from wild-type and transgenic maize endosperm. Protein bodies in the endosperm of wild-type seeds are round (*A* and *C*) compared with the misshapen protein bodies in the endosperm of seeds expressing the 24-kDa α-zein gene (*B* and *D*). Comparison of these protein bodies at lower magnification revealed that they appear as discreet spheres (circles in cross-section) in cells not expressing the 24-kDa α-zein gene (*A*), whereas they have a convoluted shape and are aggregated in cells expressing this gene (*B*). Comparison of these protein bodies at higher magnification highlights their differences in the internal structure. In cells not expressing the gene, the darkly staining zein proteins are found primarily on the periphery of the protein bodies (*C*, arrowheads), whereas in cells expressing the gene, the protein bodies have many locules of darkly staining zein in their interior (*D*, arrowheads). The irregularities seen in *B* and *D* are similar to those found in protein bodies of W64A*fl2*. PB, protein body; RER, rough endoplasmic reticulum; S, starch grain. Bars = 0.5 μm.

associated with the maize *fl2* mutation (16, 17). The experiments described here show that expression of the mutant gene in transgenic maize plants leads to the accumulation of the 24-kDa α-zein protein, which is correlated to the overexpression of BiP and the appearance of malformed protein bodies. Because these characteristics of the seed from the transgenic plants resemble the phenotype of the *fl2* mutant, our observation conclusively demonstrates that *fl2* is a structural mutation in a 22-kDa α-zein gene.

The mechanism whereby the 24-kDa α-zein protein disrupts the normal development of the protein body, leading to an altered texture of the mature endosperm, is not known. One possible explanation is that the mutant hydrophobic α-zein protein is anchored to the ER membrane and remains on the outer surface of the protein body, thereby disrupting protein interactions that are important to the maintenance of the spherical shape. Normally, α-zein proteins become sequestered in the interior of the protein body (2). This hypothesis is supported by evidence showing that the 24-kDa α-zein protein is associated with the membrane fraction following translation and translocation in the presence of microsomes (18).

Our results provide evidence of the utility of maize transformation for the analysis of genetic mechanisms. The amount of the 24-kDa α-zein protein detected in the transgenic floury kernels and the W64A*fl2* kernel was similar, suggesting that the expression of the transgene may be comparable to that of the native gene. This result is encouraging, because it demonstrates that transgene expression in tissues of transformed maize plants can be adequately controlled by native promoters. Furthermore, the 3 kb of 5′ and 3.7 kb of 3′ noncoding sequence included in the α-zein genomic clone must be sufficient for directing appropriate temporal and spatial expression in the endosperm; however, it is not known whether the gene is transcribed in other tissues of the transgenic plants.

Traditionally, techniques for stable transfer of DNA to monocotyledonous cereals have lagged behind similar work with dicotyledonous plants (25). Transformation of dicot species usually involves gene delivery through infection with *Agrobacterium tumefaciens*, but monocots are not readily amenable to infection by this bacterium. Although Ishida *et al.* (26) reported stable maize transformants following infection by *A. tumefaciens*, most successful transformations of maize plants have used microprojectile bombardment to deliver DNA to embryogenically responsive cells from immature embryos. Since the first report of the generation of stable transgenic maize plants using the biolistic method (27), this technique has been used to introduce a number of agronomically important traits into corn, such as insect resistance (20, 28), viral resistance (29), and fructan production (30). These advances, along with our report, attest to a promising future for the use of transgenic technologies in the genetic study and agronomic improvement of maize.

We thank Hank Bass for helpful discussions at the inception of this work, Susan Martino-Catt for technical assistance, and Becky Boston for providing the BiP antibody. This research was supported by a grant from the Department of Energy to B.A.L. (DEFG0292ER 20079) and a postdoctoral fellowship from the National Institutes of Health to C.E.C.

1. Thompson, G. A. & Larkins, B. A. (1989) *BioEssays* **10**, 108–113.
2. Lending, C. R. & Larkins, B. A. (1989) *Plant Cell* **1**, 1011–1023.
3. Esen, A. & Stetler, D. A. (1992) *Am. J. Bot.* **79**, 243–248.
4. Mertz, E. T., Bates, L. S. & Nelson, O. E. (1964) *Science* **145**, 279–280.
5. Nelson, O. E., Mertz, E. T. & Bates, L. S. (1965) *Science* **150**, 1469–1470.
6. Hartings, H., Maddaloni, M., Lazzaroni, N., Di Fonzo, N., Motto, M., Salamini, F. & Thompson, R. D. (1989) *EMBO J.* **8**, 2795–2801.
7. Kodrzycki, R., Boston, R. S. & Larkins, B. A. (1989) *Plant Cell* **1**, 105–114.
8. Schmidt, R. J., Burr, F. A., Aukerman, M. J. & Burr, B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 46–50.
9. Geetha, K. B., Lending, C. R., Lopes, M. A., Wallace, J. C. & Larkins, B. A. (1991) *Plant Cell* **3**, 1207–1219.
10. Soave, C., Dossena, S., Lorenzoni, C., Di Fonzo, N. & Salamini, F. (1978) *Maydica* **23**, 145–152.
11. Lending, C. R. & Larkins, B. A. (1992) *Protoplasma* **171**, 123–133.
12. Fontes, E. B. P., Shank, B. B., Wrobel, R. L., Moose, S. P., O'Brian, G. R., Wurtzel, E. T. & Boston, R. S. (1991) *Plant Cell* **3**, 483–496.
13. Boston, R. S., Fontes, E. B. P., Shank, B. B. & Wrobel, R. L. (1991) *Plant Cell* **3**, 497–505.
14. Marocco, A., Santucci, A., Cerioli, S., Motto, M., Di Fonzo, N., Thompson, R. D. & Salamini, F. (1991) *Plant Cell* **3**, 507–515.
15. Zhang, F. & Boston, R. S. (1992) *Protoplasma* **171**, 142–152.
16. Lopes, M. A., Coleman, C. E., Kodrzycki, R., Lending, C. R. & Larkins, B. A. (1994) *Mol. Gen. Genet.* **245**, 537–547.
17. Coleman, C. E., Lopes, M. A., Gillikin, J. W., Boston, R. S. & Larkins, B. A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6828–6831.
18. Gillikin, J. W., Zhang, F., Coleman, C. E., Bass, H. W., Larkins, B. A. & Boston, R. S. (1997) *Plant Physiol.* **14**, 345–352.
19. Dennehy, B. K., Petersen, W. L., Ford-Stantino, C. G., Pajeau, M. & Armstrong, C. L. (1994) *Plant Cell Tissue Organ Cult.* **36**, 1–7.
20. Koziel, M. G., Beland, G. L., Bowman, C., Carozzi, N. B., Crenshaw, R., *et al.*, (1993) *Bio/Technology* **11**, 194–200.
21. Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. & Allard, R. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 8014–8018.
22. Wallace, J. C., Lopes, M. A., Paiva, E. & Larkins, B. A. (1990) *Plant Physiol.* **92**, 191–196.
23. Lending, C. R., Kriz, A. L., Larkins, B. A. & Bracker, C. E. (1988) *Protoplasma* **143**, 51–62.
24. Knecht, D. A. & Dimond, R. L. (1984) *Anal. Biochem.* **136**, 180–184.
25. McElroy, D. & Brettell, R. I. S. (1994) *Trends Biotech.* **12**, 62–68.
26. Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T. & Kumashiro, T. (1996) *Nat. Biotech.* **14**, 745–750.
27. Gordon-Kamm, W. J., Spencer, T. M., Mangano, M. L., Adams, T. R., Daines, R. J., Start, W. G., O'Brien, J. V., Chambers, S. A., Adams, W. R., Willetts, N. G., Rice, T. B., Mackey, C. J., Krueger, R. W., Kausch, A. P. & Lemaux, P. G. (1990) *Plant Cell* **2**, 603–618.
28. Armstrong, C. L., Parker, G. B., Pershing, J. C., Brown, S. M., Sanders, P. R., *et al.*, (1995) *Crop Sci.* **35**, 550–557.
29. Murry, L. E., Elliott, L. G., Capitant, S. A., West, J. A., Hanson, K. K., Scarafia, L., Johnston, S., DeLuca-Flaherty, C., Nichols, S., Cunana, D., Dietrich, P. S., Mettler, I. J., Dewald, S., Warnick, D. A., Rhodes, C., Sinibladi, R. M. & Brunke, K. J. (1993) *Bio/Technology* **11**, 1559–1564.
30. Caimi, P. G., McCole, L. M., Klein, T. M. & Kerr, P. S. (1996) *Plant Physiol.* **110**, 355–363.

JOURNAL OF VIROLOGY, Feb. 2002, p. 1339–1348
0022-538X/02/$04.00+0 DOI: 10.1128/JVI.76.3.1339–1348.2002

Vol. 76, No. 3

Analyses of Genotypic Diversity among North, South, and Central American Isolates of *Sugarcane Yellow Leaf Virus*: Evidence for Colombian Origins and for Intraspecific Spatial Phylogenetic Variation

Francis Moonan and T. Erik Mirkov*

Department of Plant Pathology and Microbiology, Texas A&M University System Agricultural Experiment Station, Weslaco, Texas 78596

Received 8 May 2001/Accepted 23 October 2001

We have analyzed the genotypic diversity of *Sugarcane yellow leaf virus* (SCYLV) collected from North, South, and Central America by fingerprinting assays and selective cDNA cloning and sequencing. One group of isolates from Colombia, designated the G-population, has been identified as residing at the root node between a separable superpopulation structure of SCYLV and other members of the family *Luteoviridae*, indicating that the progenitor viruses of the North, South, and Central American isolates of the SCYLV superpopulation most likely arose from a C-population structure. From a model of intrafamilial evolution (F. Moonan et al., Virology 269:156–171, 2000), a prediction could be made that within the SCYLV species, the capacity of genomic sequence divergence would range from lowest in the capsid protein open reading frame 3 (ORF 3) to highest in a region spanning across the carboxy-terminal end of the RNA-dependent RNA polymerase ORF. We have demonstrated the validity and applicability of this intrafamilial model for the prediction of intraspecies SCYLV diversity. Analysis of spatial phylogenetic variation (SPV) within the SCYLV isolates could not be assessed by application of a "partial likelihoods assessed through optimization" (PLATO)-derived intraspecies model alone. However, application of a PLATO-derived intrafamilial model with the intraspecies-derived model allowed distinction of three forms of SPV. Two of the SPV forms identified correspond to the extremes in a continuum of sequence evolution displayed in a SCYLV superpopulation structure, and the third form was diagnostic of a C-population structure. The application of these types of models has value in terms of predicting the types of SCYLV intraspecies diversity that may exist worldwide, and in general, may be useful in application for more informed design of transgenes for use in the elicitation of homology-dependent virus resistance mechanisms in transgenic plants.

In the Americas, *Sugarcane yellow leaf virus* (SCYLV) is associated with a disease referred to as yellow leaf syndrome (YLS [3, 5, 6, 47, 55]), although a similar disease via phytoplasma infection may produce otherwise identical symptomology and is also currently referred to as YLS disease in certain areas of the world (6, 45). In sugarcane, losses of as high as 50% have been estimated to have occurred in field sites as the result of virus-induced YLS (55).

The complete SCYLV genome has been sequenced and characterized and most closely resembles viruses of the family *Luteoviridae* in the genus *Polerovirus* (36, 48). The results of these studies indicate that SCYLV, like *Polerovirus* members of the family *Luteoviridae*, encode at least six definable open reading frames (ORFs) typically listed as ORFs 0 to 5 (4, 7, 31, 32, 35, 36, 43, 48). ORFs 0 and 1 are thought to produce peptides via alternate translational start sites from the monopartite genomic RNA of SCYLV, and ORFs 3 and 4 are thought to produce peptides via alternate translation from a subgenomic RNA. These studies also indicate that a fusion peptide comprising a sequence encoded by ORF 2 is produced by a −1 frameshift during ORF 1 translation and that a fusion peptide comprising ORF 5 is produced via a translational readthrough of the peptide encoded by ORF 3, a finding consistent with the genome organization of well-characterized members of the genus *Polerovirus* (4, 7, 31, 32, 35, 36, 43, 48). The SCYLV peptides encoded by ORFs 0 and 1 are of unknown function and a protease, respectively, and the peptides encoded by ORFs 3, 4, and 5 appear to be structural proteins comprising the virion particle, with the ORF 3 sequence the primary capsid protein (36, 48). The peptide sequence encoded by ORF 2 appears to be multifunctional, including sequence for both an RNA-dependent RNA polymerase (RdRp) and a putative genome linked viral protein (VPg), and the VPg peptide is thought to be processed from this multifunctional peptide by proteolytic cleavage and covalently conjugated to the 5′ terminus of the SCYLV genome, similar to that in members of the genus *Polerovirus* (4, 7, 31, 32, 35, 36, 43, 48).

SCYLV as defined by the Seventh Report of the International Committee on the Taxonomy of Viruses (54) is an unclassified member of the family *Luteoviridae*, although its overall genome structure most closely resembles that of members of the genus *Polerovirus* (36, 48). The *Luteoviridae*, as currently defined, include the genera *Luteovirus*, *Polerovirus*, and *Enamovirus* (54). The structure and classification of the sequences encoding the RNA-dependent RNA polymerase of the *Luteoviridae* have served as the primary basis for generic distinction

* Corresponding author. Mailing address: Department of Plant Pathology and Microbiology, Texas A&M University System Agricultural Experiment Station, 2415 East Highway 83, Weslaco, TX 78596. Phone: (956) 968-5585. Fax: (956) 969-5260. E-mail: e-mirkov@tamu.edu.

TABLE 1. Origins of North, South, and Central American isolates of SCYLV and nature of the genotypic data collected for this study[a]

Accession	Sugarcane variety	Geographic origin	Primary genotyping source(s)	GenBank accession no.
T6	CP65-357	Weslaco, Tex.	Complete genome SCYLV-A	AF157029
F1	CP65-357	Canal Pt., Fla.	Complete genome SCYLV-F	AJ249447
F2	CP72-2086	Clewiston, Fla.	CAPS markers of amplicons	NA
L1	LHo83-153	Baton Rouge, La.	PROREP = pFM668(L1-1), pFM669(L1-2); REPUTR = pFM676(L1-1), pFM677(L1-2); CPRT = pFM623(L1)	AF369923
B0	SP71-6163	Campinas, São Paulo, Brazil	Partial cDNA sequences	AF141385, AF160474
B1	SP71-6163	Piracicaba, São Paulo, Brazil	PROREP = pFM630(B1-1), pFM631(B1-2); REPUTR = pFM708(B1-1), pFM709(B1-2); CPRT = pFM629(B1)	AF369925
B2	SP71-1406	Piracicaba, San Paolo, Brazil	CAPS markers of amplicons	NA
N6	Q136	Santa Rosa, Argentina	PROREP = pFM688(N6-1), pFM689(N6-2); REPUTR = pFM686(N6-1), pFM687(N6-2); CPRT = pFM675(N6-1), pFM674(N6-2)	AF369926
G2	CP92-1654	Santa Lucia, Guatemala	PROREP = pFM693(G2-1), pFM692(G2-2); REPUTR = pFM702(G2-1), pFM703(G2-2); CPRT = pFM696(G2-1), pFM697(G2-2)	AF369924
G3	Q50	Santa Lucia, Guatemala	CAPS markers of amplicons	NA
G6	PR7619	Santa Lucia, Guatemala	CAPS markers of amplicons	NA
G7	NA56-75	Santa Lucia, Guatemala	CAPS markers of amplicons	NA
G8	Ragnar	Santa Lucia, Guatemala	CAPS markers of amplicons	NA
G9	PR1013	Santa Lucia, Guatemala	CAPS markers of amplicons	NA
C1	SP71-6163	Cali, Colombia	PROREP = pFM715(C1-1), pFM716(C1-2); REPUTR = pFM704(C1-1), pFM705(C1-2); CPRT = pFM719(C1-1); pFM720(C1-2)	AF369927
C3	CC84-75	Cali, Colombia	PROREP = pFM608(C3-1), pFM609(C3-2); REPUTR = pFM724(C3-1), pFM725(C3-2); CPRT = pFM775(C3-1), pFM773(C3-2)	AF369928
C4	CC85-964	Cali, Colombia	PROREP = pFM670(C4-1), pFM671(C4-2); REPUTR = pFM708(C4-1), pFM709(C4-2); CPRT = pFM624(C4)	AF369929

[a] GenBank accession numbers corresponding to the T6 (SCYLV-A), F1 (SCYLV-F), and B0 isolates have respectively been reported by Moonan et al. (36), Smith et al. (48), and Maia et al. (29). The remaining GenBank accession numbers listed in the table are newly reported here. The accession designation labels in the table are used in Fig. 1A-C, 2, 3, and 4, and the designations in parentheses listed under the genotyping source portion of the table were used to derive the paired data sets of deduced peptide sequences analyzed in Fig. 1D and E from the indicated pFM plasmid clones.

within the family (7, 54). Based upon the classification system of Koonin and Dolja (25), *Luteovirus* members have genomes that encode RNA-dependent RNA polymerases classified as supergroup II; *Polerovirus* members have RdRps classified as supergroup I; and the sole member of the *Enamovirus* genus, *Pea enation mosaic virus* 1 (PEMV-1), has an RdRp sequence that contains sequences of both RdRp supergroup I and II origin (36). The genomes of PEMV-1, *Soybean dwarf virus*, and SCYLV exhibit spatial phylogenetic variation (SPV [18]) that is thought to have arisen via recombination between polerovirus and luteovirus ancestors after the divergence of these two progenitor groups (4, 7, 17, 18, 31, 32, 35, 36, 43, 54).

Our previous comparisons of the SCYLV genome with other members of the family *Luteoviridae* involved the development of an intrafamilial model of SPV (36). Based upon this model, we predicted that the SCYLV genomic sequence diversity would be lowest in a region spanning the capsid protein ORF 3, low to intermediate in a region spanning the protease ORF 1 to within the RdRp-encoded ORF 2, and highest in a region spanning from the RdRp to an untranslated sequence located 5′ proximal to an ORF 3 and 4 polycistron. To test the predictions of this model, we collected SCYLV isolates from field sites throughout North, South, and Central America and analyzed the genotypic diversity among these isolates. The results of our analyses not only demonstrate the validity and applicability of this model, in terms of predicting the types of SCYLV intraspecific diversity, but have also aided us in identifying a population of isolates from Colombia, which we refer to as the C-population, that most likely represents the ancestral population of all other sampled American SCYLV isolates.

MATERIALS AND METHODS

Plant materials, SCYLV infectivity diagnosis, and amplicon production. Plant tissues samples were collected from sugarcane plants with leaf-yellowing symptoms diagnostic of YLS. RNA extraction and purification from sugarcane tissue samples for Northern blot analyses were performed essentially as described by Ingelbrecht et al. (23) by using a double-stranded DNA probe derived from pFM261 (36). A summary of the samples collected that were positive for SCYLV infection is presented in Table 1. Amplicons were produced as diagnostic fragments by using reverse transcription (RT)-PCR methods and were designated PROREP (protease to replicase), REPUTR (replicase to untranslated region and beginning of capsid protein), and CPRT (capsid protein to beginning of capsid readthrough protein region). The positions of the PROREP, REPUTR, and CPRT amplicons, in relation to the SCYLV-A genome, are shown in Fig. 4. SCYLV first-strand cDNA was produced by RT of 3 μg of total plant RNA with a Universal Riboclone cDNA Synthesis System (Promega, Palo Alto, Calif.), in a 25-μl volume at 42°C and with 5 U of avian myeloblastosis virus reverse transcriptase and 50 ng of oligonucleotide primers, according to the manufacturer's instructions. The following oligonucleotide primers were used for the RT reactions: oFM359 (5′-GCTCTCCACAAAGCTATCT-3′), oFM387 (5′-CTGACATTCCTTCGTGAGC-3′) and oFM361 (5′-TGTTTTCACGATGTGGTTC-3′). RT reaction mixtures were diluted with an additional 12.5 μl of water and then utilized in subsequent PCRs. For each accession, PCRs were done in triplicate for each of the PROREP, REPUTR, and CPRT amplicons. PCRs were done in 50-μl reaction volumes of 1× Qiagen *Taq* Polymerase Buffer with $MgCl_2$, 2 mM deoxynucleoside triphosphates, and 1 to 5 μl of first-strand cDNA

mixture. PCRs were for 30 cycles of 95°C for 1 min, 52°C (PRO-REP) or 58°C (REPUTR and CPRT) for 2 min, and 72°C for 2 min. PCRs were done to produce the following amplicons with the following primer combinations: PRO-REP with oFM359 and oFM323 (5'-CAGACATTGCTGATTAC-3'), REPUTR with oFM387 and oFM386 (5'-AGATAGCTTTGTGGAGAGC-3'), and CPRT with oFM361 and oFM366 (5'-GCTCACGAAGGAATGTCAG-3'). Amplicons were size fractionated on agarose gels, and amplicons were gel purified with a Geneclean II Kit (Bio 101, Carlsbad, Calif.) according to the manufacturer's instructions prior to their use in fingerprinting or cloning.

Production and analysis of CAPS fingerprint gels. Amplicons generated in triplicate, from independent PCRs, were used for DNA fingerprinting. Gel-purified amplicon fragments were digested with both *Sau*3AI and *Taq*I restriction enzymes to generate the SCYLV-cleaved amplified polymorphic sequence (CAPS), and these were radiolabeled by 5' overhang end-filling reactions with Klenow enzyme (Promega) and dATP, dCTP, dTTP, and [α-^{32}P]dCTP. CAPS reactions were fractionated on 6% polyacrylamide sequencing gels under standard denaturing conditions (44). Dephosphorylated *Hin*fI-digested ϕX174 DNA was labeled with [γ-^{32}P]ATP and polynucleotide kinase and used as size markers. A total of 65 CAPS allelic fragments were scored and ranged between 27 and 384 nucleotides (nt) in length: 17 CAPS allelic fragments for the CPRT amplicons, 22 for the REPUTR amplicons, and 26 for the PROREP amplicons.

cDNA cloning, sequencing, and computational analyses. CAPS markers scored as binary data were analyzed by unweighted pair group method with arithmetic mean clustering methodology with Paup, version 4.0 beta 4 (51), and the haploid data analysis components of Popgene (57). Analyses of Nei's (37) gene diversity (h) and Nei's (39) pairwise genetic distances were calculated with Popgene and analyzed individually, as a whole, and as geographic groupings. Nucleotide and deduced peptide alignments and CAPS-associated files are deposited at Treebase (www.herbaria.harvard.edu/treebase/index.html). REPPRO, REPUTR, and CPRT amplicons derived from one or two of the three available independent PCRs were ligated into pCR4-TOPO (Invitrogen) to generate independent clones for each of the amplicons. Only one of the CPRT amplicons of the L1, B1, and C4 accessions was sequenced (Table 1, pFM623, pFM629, and pFM624). All other amplicons were cloned and sequenced in duplicate. The consensus sequences of the SCYLV isolates are deposited in GenBank and are listed in Table 1. The T6 accession represents the source material for SCYLV-A (36). Source material for SCYLV-F (48) from variety CP65-357 was obtained and designated the F1 isolate of this study. The sequence from a Brazilian isolate described by Maia et al. (29) was designated the B0 isolate. Outgroup comparisons, nucleotide alignments of the SCYLV data set with other species of the *Luteoviridae*, and the description of the origins of these genomic sequences are reported in the study by Moonan et al. (36). DNA sequence generation was performed on a contract basis by the DNA Sequencing Facility at Iowa State University, Ames. Plasmid insert sequences were assembled with Seqman II (DNAstar, Inc., Madison, Wis.). Nucleotide multiple sequence alignments were analyzed with Paup 4.0 beta 4 or beta 8 (51), MEGA 2.0 (26), PHYLIP 3.57 (15), and the neighbor-joining (NJ) method of CLUSTAL X (52, 53). Deduced peptide sequences were analyzed with the NJ method of CLUSTAL X and the quartet maximum-likelihood-based method of PUZZLE (50). PUZZLE analyses were done with both the Dayhoff and Jones, Taylor, and Thornton phylogenetic models with 1,000 quartets for the partial deduced peptide sequences of ORFs 1 and 5, as well as ORFs 3 and 4, and 10,000 quartets for ORF 2. For phylogenetic analysis of nucleotide sequences with Splitstree 2.4 (22), distance data was generated with the Hasegawa-Kishino-Yano phylogenetic model (20) of Paup 4.0 beta 8. Analyses for potential recombination sites were done by "likelihood analysis of recombination in DNA" (LARD [21]) and the Recombination in DNA Program (RDP [30]). Nucleotide sequences were analyzed for SPV by the PLATO (18) and DNAml (15) maximum-likelihood methods and the NJ method of CLUSTAL X. Because the consensus sequence for SCYLV-A contains three alternate nucleotide substitutions in the region spanning the PRO-REP amplicon-spanning region and six alternate nucleotide substitution patterns in the REPUTR amplicon-spanning region, the 8 and 64 sequentially derived sequences for SCYLV-A were initially analyzed with the 14 individual sequences of the other isolates assayed. The results (not shown) indicated that there was no overlap of the different SCYLV-A combinations with the sequences of the other isolates. By using composited pairs of the #1 and #8 PROREP and #1 and #64 REPUTR-spanning regions, which represented all individual nucleotide substitutions in SCYLV-A, two 2,835-nt SCYLV-A sequences were generated and referenced as T6-1 and T6-2. The transition/transversion ratio used in both the PLATO and DNAml was calculated with 14 generated composite 2,835-nt clonal sequences for each of the B1, N6, L1, G2, C1, C3, and C4 accessions, along with the T6-1 and T6-2 sequences. The overall transition/transversion ratio for this sequence range was calculated to be 1.5182, and this ratio was used with the DNAml (15) and the HKY phylogenetic model with PLATO.

RESULTS

Of 35 accessions collected, 16 samples tested positive for SCYLV infection and are listed in Table 1. Accessions T3, T4, and T5 (not listed) functioned as negative controls and represent plants independently generated via meristem culture from the T6 accession via previously described methods (46). Numerous anecdotal reports suggest that SCYLV is probably distributed worldwide. Our survey of the Americas confirms the previous reports that SCYLV is endemic in all of the major sugarcane-growing regions in the continental United States (5, 36, 45, 48), as well as in Brazil (29, 55), but also indicates that the geographic range of this virus may also now be formally expanded to include Guatemala, Colombia, and Argentina.

Fingerprint analyses of isolates. A total of 65 CAPS allelic fragments were scored in binary from the amplicons sampled, and genotypic clustering was performed with the UPGMA methodology of Paup. The resulting dendrogram is shown in Fig. 1A. Fingerprint patterns for the T6, F1, F2, G3, G6, G7, G8, and G9 amplicons were identical in triplicate, while fingerprint patterns for the B2 and N6 isolates were identical in triplicate, and these are represented in the dendrogram in Fig. 1A as a vertical bar subtended by concave branches. An identical fingerprint patterning in triplicate implies an effective Nei (37, 38, 39) pairwise genetic distance of zero, indicating that groups with identical fingerprints are most likely representative of the same genotype. The Cali, Colombia, isolate members C1, C3, and C4, which resolved as a putative cluster in the PAUP-derived UPGMA dendrogram in Fig. 1A, exhibited differences in gene diversity as a group in comparison to the other isolates. The overall gene diversity for all of the isolates was calculated as $h = 0.0471$. In contrast, the C1/C3/C4 group gene diversity was calculated as $h = 0.0547$, while the remaining isolates were calculated as $h = 0.00248$, indicating that the bulk of the gene diversity from the samples assayed was represented primarily by the members of the C1/C3/C4 group of isolates. Analyses of Nei (39) pairwise genetic distance ranges within the C1/C3/C4 group and within the remaining isolates were 0.0313 to 0.1313 and 0 to 0.0473, respectively, which reflect this difference in calculated gene diversity measurements.

Sequence analyses of isolates. Based upon the analysis of the genotypes represented in Fig. 1A, amplicons for B1, N6, L1, G2, C1, C3, and C4 were selected for DNA sequencing. From the five or six independent amplicons cloned and sequenced for each accession, which represented paired data sets, a 2,832- or 2,835-nt consensus sequence for each isolate was then derived, which corresponded to nt 1518 to 4352 of the SCYLV-A and T6 isolate genome (36). Alignment of these data sets indicated that one predominant insert/deletion (INDEL) could be identified in the alignment, which corresponded to one gap of 3 nt, in the six individual paired data sets from the REPUTR amplicons of the C1, C3, and C4 isolates. This INDEL corresponded to the nt 3512 to 3514 of the SCYLV-A genome, which is positioned in an untranslated region of the genome, between ORF 2 and ORFs 3. Pairwise genetic distances were calculated with the Jukes-Cantor (J-C) method of MEGA 2.0 with 1,000 replicates, and mean sequence divergence measure-

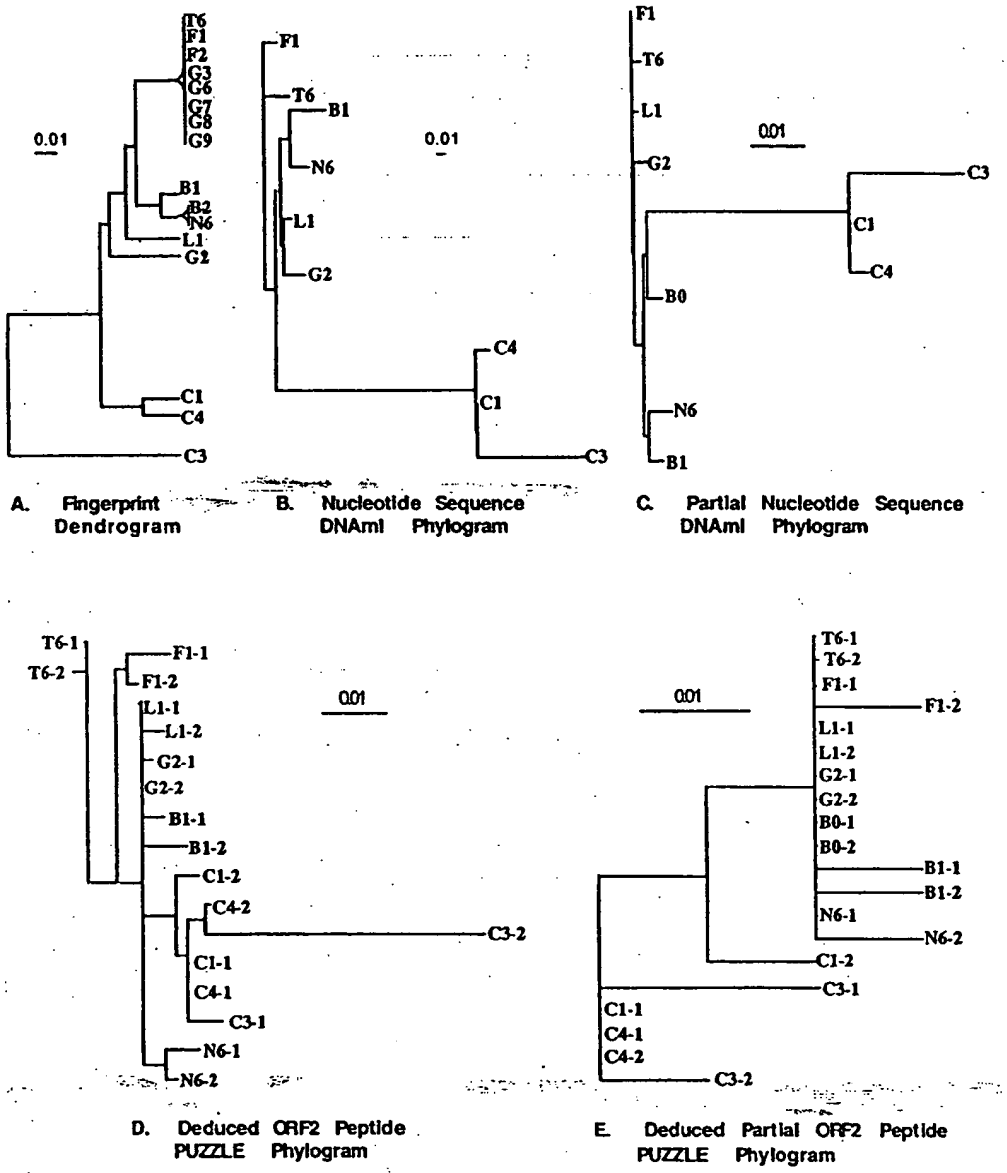


FIG. 1. Phylogenetic relationships assessed with both a UPGMA dendrogram derived from fingerprinting data (A) and phylograms derived from nucleotide (B and C) and deduced peptide (D and E) data. The designations of the isolates and the origins of the sequence or fingerprint data used in the diagrams are given in Table 1. Nucleotide sequence alignments were analyzed with the F84 phylogenetic model of DNAml of the PHYLIP package, and deduced peptide sequence alignments were analyzed with the maximum-likelihood based quartet puzzling method of PUZZLE. The nucleotide sequence ranges analyzed for the isolate phylograms (B and C), respectively, correspond to nt 1518 to 4352 and nt 3144 to 4238 of the SCYLV-A genome described by Moonan et al. (36). Relationships between the deduced peptide sequences of RdRp ORF 2 of SCYLV are shown for both the complete RdRp peptide sequences (D) and a partial RdRp sequence corresponding to amino acid residue positions 471 to 572 of the RdRp of SCYLV-A (E).

ments were calculated. The MEGA 2.0-calculated mean sequence diversity for the set of nine consensus sequences, excluding the B0 data, was 0.01835, with a standard error (SE) of 0.0166; the mean diversity of the C1/C3/C4 and T6/F1/L1/G2/B1/N6 sequences as groups was 0.00557 (SE = 0.0084); and between these same groups the nucleotide diversity was 0.01278 (SE = 0.00159). The calculated F84 and J-C pairwise genetic distances within the C1, C3, and C4 isolates, respectively, ranged from 0 0.0018 to 0.0440 and 0.0018 to 0.0160, and within the T6, F1, L1, G2, B1, and N6 isolates they ranged from 0.0036 to 0.0107 and 0.0036 to 0.0107.

Both the NJ method of CLUSTAL X and the maximum-likelihood method of DNAml were used to generate phylograms. These were compared to the trees generated by the paired sample sets. In either of these paired sample sets, or the consensus sequences, there was little difference in the tree topologies in the most common and most likely trees generated by these methods. Figure 1B shows the resulting DNAml phy-

logram with the consensus sequence data, which may be compared to the fingerprint derived data shown in Fig. 1A. A comparison of the dendrogram in Fig. 1A and the phylogram in Fig. 1B indicates that the results of the fingerprint and sequence analysis demonstrate the same pattern of similarities between isolates, supporting a contention that the C1/C3/C4 group of isolates sampled represents a population structure separable from the T6/F1/F2/L1/G2/B1/B2/N6 group of isolates sampled.

Partial sequences from a isolate of SCYLV from Campinas, São Paulo, Brazil, listed in Table 1 as a B0 isolate were analyzed in a similar fashion and represented nt 3144 to 4238 of the SCYLV-A (T6 isolate) genome. These B0 data, produced by Maia et al. (29), included the sequence spanning a small portion of SCYLV ORF 2 and the complete ORFs 3 and 4 of SCYLV, which are encoded in a polycistron within the SCYLV genome. The generated alignments indicated an INDEL of a single A nucleotide corresponding to a nucleotide position between nt 3623 and nt 3624 of the SCYLV-A (T6 isolate) genome, which is positioned in the untranslated region between ORF 2 and ORF 3. The B0 resulting DNAml phylogram for the SCYLV-A corresponding nucleotide range from 3144 to 4238 is shown in Fig. 1C. As shown in Fig. 1C, the C1, C3, and C4 isolates from this analysis resolved as a clade most closely associated with the B0 isolate. Based upon both the fingerprint and sequence-derived analyses, the Cali, Colombia, C1/C3/C4 sampled group could be considered as a separable and distinct geographic population set from the sampled T6/F1/F2/L1/G2/B1/B2/N6 geographic population set. Arbitrarily, we assigned the designation of C-population to the C1/C3/C4 population set to reflect its Colombian origins and the designation "superpopulation" to the T6/F1/F2/L1/G2/B1/B2/N6 population set to reflect its larger size from the pool of the overall samples of isolates studied, the apparent genetically contiguous characteristics of the isolates sampled and analyzed, as well as the widespread geographic distribution of the isolates sampled and studied, which spanned across geographic locales in both North and South America.

To determine the degree to which the differences in nucleotide sequence corresponded to differences in deduced peptide sequence composition of the complete and partial ORF sequences encoded, we compared the deduced peptide sequences of partial ORFs 1 and 5 and the complete ORFs 2, 3 and 4 with alignments in which the B0 isolate derived data was excluded (Fig. 1D) or included (Fig. 1E). Trees were generated by the NJ method, as well as by the maximum-likelihood method of PUZZLE. For the PUZZLE-generated trees, both the Dayhoff and JTT phylogenetic models were used, but there was no significant difference in tree topology seen with these two methods. The trees from the ORF 3 and 4 deduced peptide data, in which the B0 sequences were included or excluded exhibited no significant differences, with the SE of the branchtree lengths overlapping, thus represented the tree topology of a single clade for all of these assessments (data not shown). The trees generated by analyses of the deduced peptide sequence of ORF 2, however, consistent with our pairwise comparisons, indicated that a separate clade composed of sequences derived from the C1, C3, and C4 isolates could be delimited from the remaining isolate sequences. Figure 1D shows the PUZZLE-derived phylogram from the B0 excluded

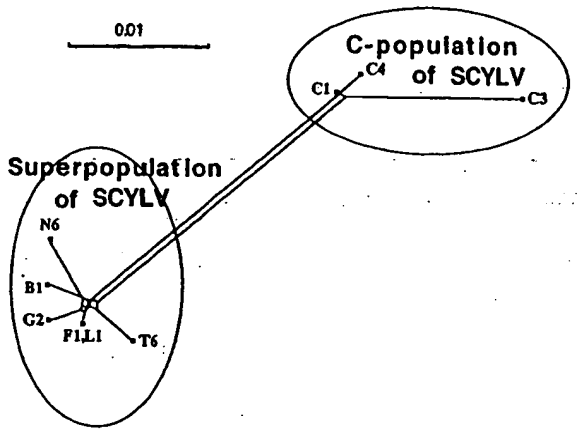


FIG. 2. Splitstree network diagram derived from nt 1518 to 4352 of the SCYLV-A genome, illustrating the most likely possible phylogenetic relationships between isolates of SCYLV. The network was produced in Splitstree from an HKY phylogenetic model produced with PAUP and corresponds to the data used to produce the DNAml phylogram illustrated in Fig. 1B. The designations of the isolates and the origins of the sequences used to produce the network are given in Table 1.

alignment, and Fig. 1E shows the results from inclusion of the B0 isolate data. As shown in Fig. 1D and E, the individual deduced peptide sequences of the six C1, C3, and C4 sampled sequences resolved in phylograms as a clade separable from the remaining sequences, as did the phylograms generated by the nucleotide sequence-generated phylograms (Fig. 1B and C).

The isolates' pairwise genetic distances were calculated with the same nucleotide sequence alignments used to produce Fig. 1B and 1C, with the HKY phylogenetic model of Paup 4.0 beta 8, and the resulting data were further analyzed with Splitstree 2.4. As shown in Fig. 2, Splitstree output from the B0 excluded alignment, which represented the nucleotide range from nt 1518 to nt 4352 of the SCYLV-A genome, produced a network diagram in which a series of short quadrangles are generated that extend from the T6 isolate to the N6 isolate. This quadrangle set is linked to a long quadrangle, which at the extreme is associated with the C1, C3, and C4 isolate sequences. In the three short quadrangles represented in Fig. 2, the T6 isolate is positioned from one quadrangle which is opposed by a B1/N6-associated quadrangle, in which a third quadrangle may be joined to these two, to which the G2/F1/L1 isolates are associated. The C1 isolate is placed in this network within the quadrangle split that leads directly to a structure of three short quadrangles within the SCYLV superpopulation. Splitstree output from a B0 inclusive alignment which represented nucleotides ranging from nt 3144 to nt 4238 of the SCYLV-A genome also produced a network diagram similar to that shown in Fig. 2 (data not shown). Although the C-population cluster from this B0 inclusive output was also anchored at a node at which the C1 isolate was placed, rather than a quadrangle extending from the C-population to the superpopulation cluster, a single branch linked the two. In addition, rather than a set of networked quadrangles at the superpopulation basal node, the superpopulation isolates were distributed in a

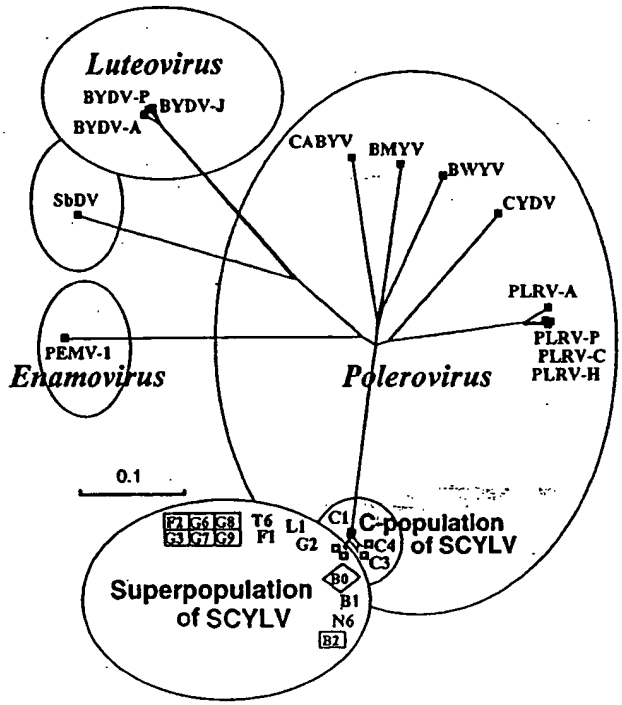


FIG. 3. Splitstree network diagram derived from nt 1518 to 4352 of the SCYLV genome, illustrating the most likely phylogenetic relationships between the SCYLV isolates and other members of the family *Luteoviridae*. The network was produced in Splitstree from an HKY phylogenetic model produced with PAUP. The designations of the isolates and the origins of the SCYLV sequences used to produce the network are given in Table 1, and the origins of the *Luteoviridae* sequences are described in Moonan et al. (36). Isolate clusters formally designated as belonging to the *Enamovirus*, *Polerovirus*, and *Luteovirus* genera are encircled, as are the postulated superpopulation and C-population clusters of SCYLV. Within the network diagram, assignment of the B0 isolate of Maia et al. (29) (in diamond) is based on analyses of partial sequence information data shown in Fig. 1C and E, and assignment of other isolates of SCYLV to the superpopulation (in box) is based upon UPGMA analysis of fingerprint data, as shown in Fig. 1A.

radial tree topology extending from a central superpopulation node and included the B0 isolate sequence.

Sequences represented in Fig. 1B and 2 were aligned with those of other members of the family *Luteoviridae*, with an alignment that has been previously described used as a guide (36). Pairwise genetic distances were calculated with the J-C model of MEGA 2.0, the J-C and F84 phylogenetic models of DNAml, and the HKY model of Paup 4 beta 8. Data from the Paup 4.0 beta 8-produced model were analyzed with Splitstree 2.4. As shown in Fig. 3, output from Splitstree placed the C1 isolate sequence at a node of a quadrangle from which the C3 and C4 isolates branched and at a node placed directly at a branch that extends into a node from which predominantly *Polerovirus* genomic sequences were generated. This C1 isolate node, as shown in Fig. 3, also leads directly to a node from which two short branches lead to individually separable but internally unresolvable clusters of the T6/F1/L1/G2 and the B1/N6 isolate sequences.

Analyses of SPV. Using the paired data sets of 16 sequences, we analyzed the 2,835-nt sequence range of SCYLV with both LARD and RDP. LARD was initially used for detection of recombination and SPV in *Dengue virus* (21), and RDP is a program for aiding in the detecting of recombination among a set of aligned viral sequences (30). Utilizing combinations of parental sequences with both of these programs, no potential recombination sites could be identified. By using the same paired data set, we then performed an analysis of SPV in the same fashion described by Moonan et al. (36), except that the transition/transversion ratio utilized was 1.5182. From the intraspecies PLATO analysis output, z values of >3.57 were statistically significant and are shown in Fig. 4A, below the corresponding PLATO output from the intrafamilial derived range used with the SCYLV-A genome (36). NJ method results and DNAml trees were generated and analyzed with the intraspecies derived PLATO output ranges, but the corresponding trees for these ranges showed little congruence in tree topologies generated by these two phylogenetic methods. The sequence ranges were then reanalyzed by using the corresponding breakpoints extrapolated from the *Luteoviridae* sequence alignment derived intrafamilial PLATO output (36), along with the breakpoints generated from the intraspecies PLATO output. The results, shown in Fig. 4, indicate that SPV could be delimited within the SCYLV isolates by utilizing this approach. From the side-by-side individual tree comparisons derived from the NJ method and DNAml tree topologies, four tree topologies could be delimited, which are shown in Fig. 4B. Analyses of the resulting trees indicated that the overall distribution of SPV allowed the assignment of sequences to three possible groups: (i) group A, containing the T6, F1, L1, and G2 isolate sequences; (ii) group B, containing the B2 and N6 isolate sequences; and (iii) group C, containing the C1, C3, and C4 isolate sequences. These groups were delimited by the four tree topologies shown in Fig. 4B: (i) an I-tree ("I" for indistinguishable), in which there was no statistical clustering of a sequence range into any cluster beyond a single clade; (ii) a B-tree, in which B group designates clustered significantly different from the other isolates, which among themselves showed no significant statistical difference in terms of genetic distances; (iii) a C-tree, in which C group designates clustered significantly different from the other isolates, which among themselves showed no significant statistical difference in clustering; and (iv) an X-tree, in which the group A, group B, and group C isolate clusters were separated in a trichotomy.

DISCUSSION

SCYLV population structures may be delimited, and their evolutionary relationships inferred, by utilizing phylogenetic models. In terms of plant viral strain discrimination, the most common molecular criterion used is based on comparisons of nucleotide sequence identity, as well as the derived deduced peptide sequence similarities of complete ORFs. The most commonly utilized deduced peptide sequences used in the discrimination of strains encode the RdRp and capsid proteins. Our sampling of genotypic diversity of SCYLV has thus initially involved the cloning and characterization of regions of the genome spanning these ORFs, and ca. 48% of the SCYLV genome has been sampled in these analyses. In part, we col-

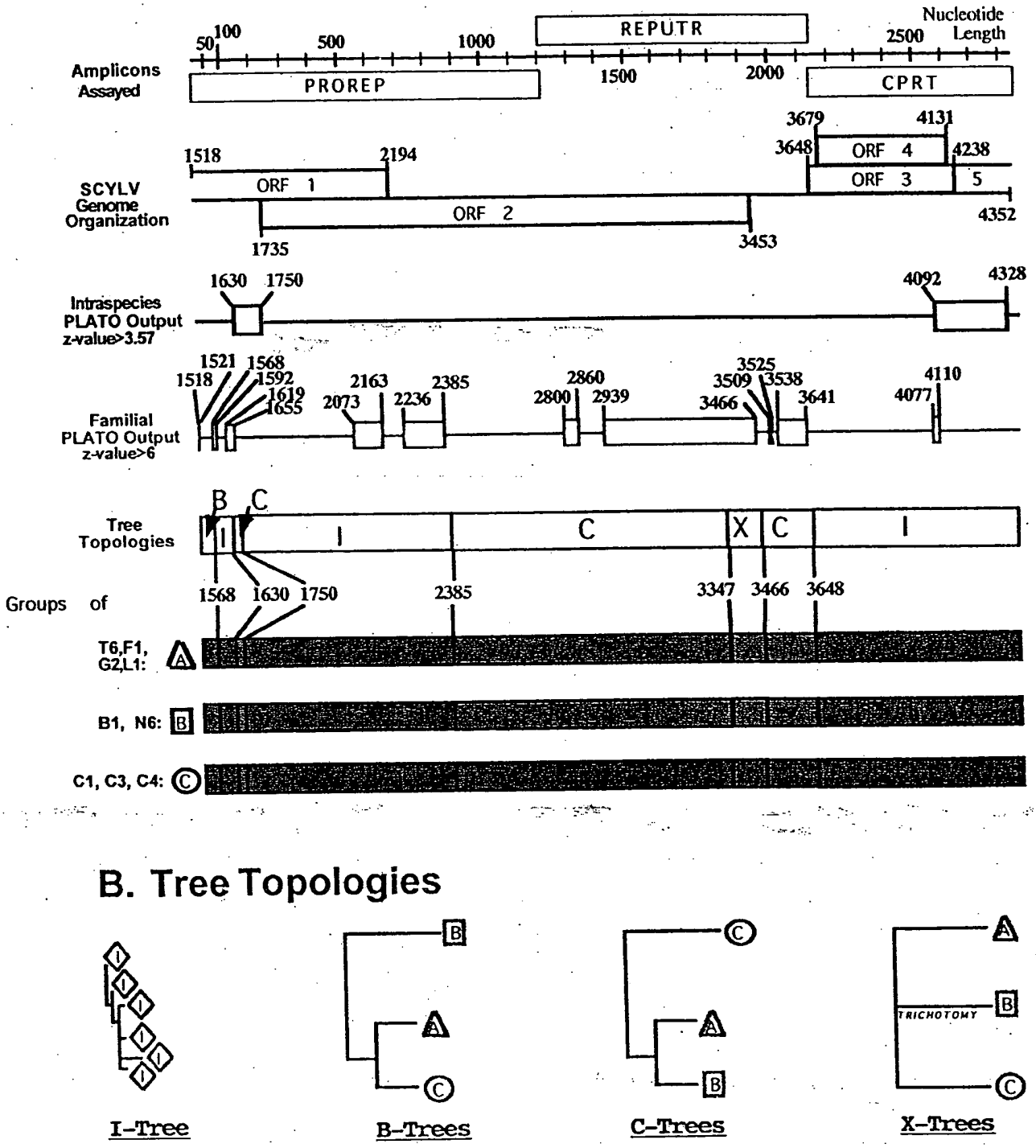
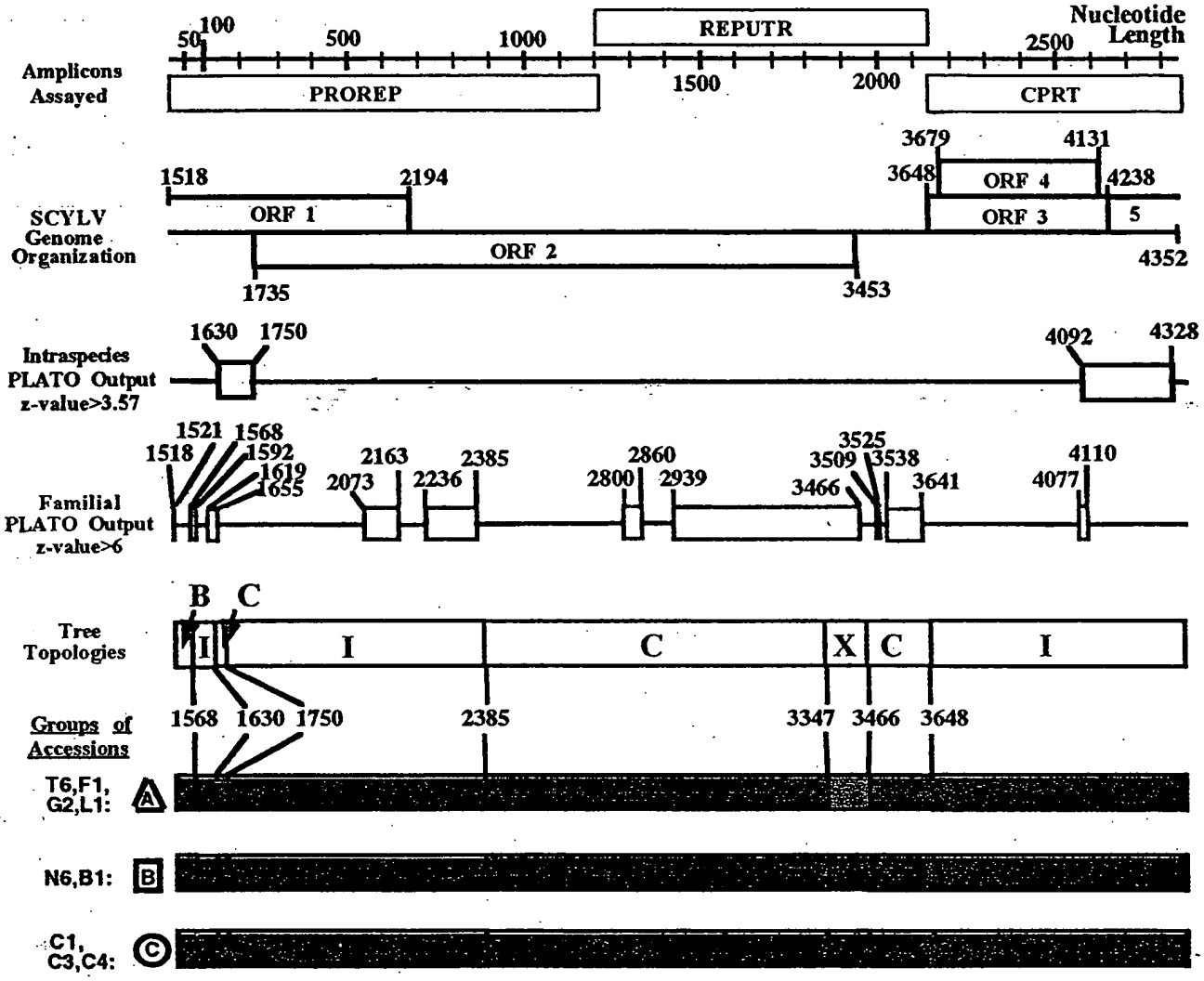
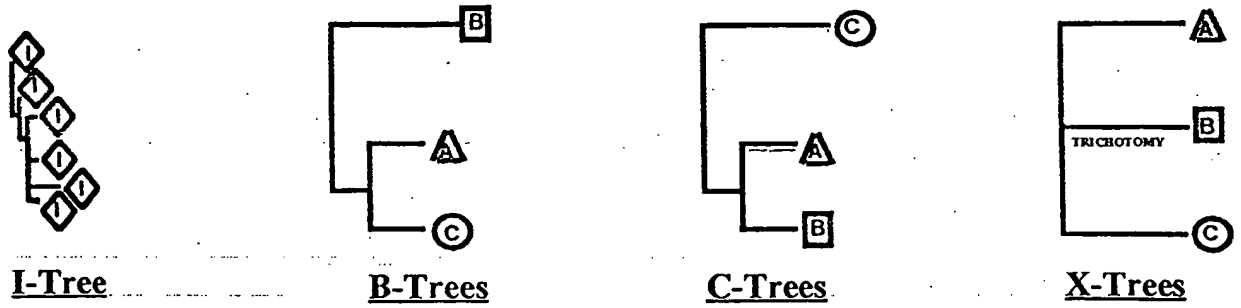


FIG. 4. Analysis of SPV in isolates of SCYLV from North, South, and Central America, corresponding to nt 1518 to 4352 of the SCYLV-A genome (A), and the tree topology types exhibited in this SPV (B). (A, top to bottom) Relative position of the clonal inserts used to derive the SCYLV sequences for this region; genomic organization of the SCYLV ORFs corresponding to this region, corresponding isolate data-derived PLATO output ranges shown as boxed ranges corresponding to the SCYLV genome; corresponding PLATO output ranges derived from an intrafamilial model; distribution of tree topologies corresponding to regions of SPV; and grouping of isolates into three groups (groups A, B, and C), corresponding to the exhibited patterns of SPV. The SCYLV-A genome used as a reference, the direct extrapolation of the PLATO intrafamilial analysis-based output, and a description of this method of SPV analysis are derived from Moonan et al. (36).

A. Assessment of Spatial Phylogenetic Variation



B. Tree Topologies



lected and analyzed SCYLV accessions from North, South, and Central America to determine whether we could discriminate, by using molecular criteria, new strains of the virus. The partial ORF 1 and 5 deduced peptide sequence comparisons revealed up to 5% amino acid sequence differences (data not shown), but interpretations from the small lengths of the deduced peptides from these partial sequences should be considered dubious. Pairwise comparisons of the deduced peptide sequences of ORF 2 indicated some differences which might be indicative of an amino acid identity difference of as high as 5% between members of the SCYLV superpopulation and the C-population. However, the estimates of SEs from the within-accession measurements and between-accession estimates indicated overlap by an SE in each of the instances, and therefore interpretations of strain distinction from our sampling are questionable. Our results suggest that isolates of the C-population represent a different strain from the SCYLV superpopulation, but resolution of this issue will require either increased sampling or the addition of pertinent biological data.

The use of phylogenies based on molecular data is a commonly accepted approach to defining population structures (16, 19, 24, 37–41), and the advantages of using phylogenies in this fashion have been described (14, 40). Migration of RNA viruses has been shown to fix classes of genomes that may exhibit variable degrees of fitness (11). In terms of RNA viruses, most migrations involve either vector transmissions or dispersal of small samples of infected organisms or virus-carrying vectors to new geographic locales (8–11). The lineage of these viruses may be traced by various means, including the identification of dispersed recombinant virus forms (18, 21). In the *Luteoviridae*, evidence exists that RNA recombination has occurred episodically at two levels. The first-level episodic RNA recombination event(s) accommodates the divergence of the two major classes of *Luteoviridae* genomes that are currently represented by the *Luteovirus* and *Polerovirus* genera: *Luteovirus* members produce an RNA-dependent RNA polymerase protein, lack a VPg protein, and have a 5′ gene organization most closely allied with members of the *Sobemovirus* genus, whereas *Polerovirus* members express a VPg, as well as an RdRp most closely allied with the RdRp proteins expressed by members of the *Dianthovirus* or *Carmovirus* genus (4, 7, 31, 32, 35, 36, 43). At a secondary level, evidence indicates that episodic recombination between polerovirus-polerovirus (17) and luteovirus-polerovirus ancestors (36, 43, 48) has occurred since the divergence of these two groups (4, 7, 31, 32, 35).

Overall, the phylogenetic analysis-based data we present in this study indicates that a majority of the isolates we studied are associated with a superpopulation structure of SCYLV. The results also indicate that the C-population of SCYLV represents a population structure from which the SCYLV superpopulation most likely emerged. Two primary hypotheses might be considered to explain the phylogenetic relationship between the SCYLV populations we refer to as the superpopulation and the C-population: (i) the superpopulation was derived from a separable and distinct C-population via either founder effects or genetic drift or (ii) the C-population was derived from the superpopulation via recombination with some unknown *Luteoviridae* family member that exists at a closer genetic distance to the main *Polerovirus* clade, to which SCYLV is most closely allied. The evidence primarily supports the first conclusion. First, our attempts to use different methods of computational technology, i.e., to determine whether the SPV we observed is derived from recombinatorial events, indicated that there was no significant evidence for detectable recombination events with the isolates sampled and analyzed. Second, a comparison of the translated ORF 2 (RdRp) gene products from the C-population in relation to the sampled nucleotide sequences indicated that the majority of those sequence changes in the C-population are silent in reference to the same homologous sites in the SCYLV superpopulation. Third, our current sampling methods indicate that the isolates that constitute the C-population have higher estimates of gene diversity than the isolates from the superpopulation, which would be a predictable observation if founder effects were involved in a hypothesized evolution of the SCYLV superpopulation from a C-population member or progenitor. The mean Nei (37) gene diversity estimates derived from fingerprint data are $h = 0.0248$ for the superpopulation set and $h = 0.0547$ for the C-population set for all of the CAPS markers. The range of genetic distances measured from the nucleotide data also supports this conclusion.

Based upon the SPV shown in Fig. 4, the dendrogram in Fig. 1A, and the Splitstree network in Fig. 3, the B0 and B2 isolates could be assigned in a classification scheme as a population subgroup, group B, which along with the B1 and N6 isolates represents one end of a derived continuum of sequence evolution within the SCYLV superpopulation structure. Based primarily on the SPV shown in Fig. 4, the T6, F1, F2, L1, G2, G3, G6, G7, G8, and G9 isolates could be considered as group A, constituting the other end of a continuum of isolates in the superpopulation. This partitioning scheme may be represented in the Splitstree network in Fig. 3 by the placement of the B0 (in diamond) and F2, G3, and G6 to G9 isolates (in box). This group A-group B partitioning scheme should be considered artificial. The positioning of the L1/G2 sequences in the Splitstree network in Fig. 2 in a quadrangle adjoining the F1/T6 and N6/B1 isolates suggests that the SCYLV superpopulation members more accurately represent a continuum of sequence evolution. The value of a group A-group B partitioning scheme is in its merits as a guide toward developing molecular epidemiologic models for the geographic migration and dispersal of SCYLV. If a parsimonious philosophy (i.e., that the lowest number of steps leading to a result is the most likely path taken to arrive at that result) is applied to this somewhat artificial group classification, a progression of SCYLV evolution in which C-population isolates have given rise to group B-type isolates, from which group A-type isolates are further evolved, is the simplest conclusion. If we utilize the group A-group B superpopulation and C-population forms of classification, our survey would indicate that group A isolates are distributed between North America and Guatemala, that group B isolates are distributed between Argentina and Brazil, and that the C-population isolates have only been identified from Colombia. None of our geographic sites yielded genotypes assignable to more than one grouping from this somewhat artificial form of classification. Although these observations may be due to a lack of extensive enough sampling, they could also be the result of founder effect involvement in the intraspecies evolution of SCYLV.

In this study, we have examined viral isolate relationships in

part with network diagrams generated by Splitstree (22). The diagrams generated by an analysis with Splitstree utilize a form of split decomposition analysis (1, 13) that is thought to provide a more reliable interpretation of phylogenetic data, because evolutionary data often contain conflicting phylogenetic data from which alternate and also highly likely phylogenetic relationships might not otherwise be illustrated (12, 13, 22, 27, 28, 33). For example, in the network diagram in Fig. 2, a series of quadrangles are linked in relation to an observed continuum of sequence evolution of a superpopulation of SCYLV that extends from the T6 isolate to the N6 isolate. Three main quadrangles are represented, of which the T6 isolate is positioned from one quadrangle which is opposed by a B1/N6 associated quadrangle, and in which a third quadrangle may be joined to these two, in which the G2/F1/L1 isolates are associated. Figure 2 also shows that although the C1 isolate may be placed within a quadrangle node that leads to the central quadrangles of the superpopulation structure, effectively placed at what is presumed to be the shortest genetic distance and therefore most likely intersection with this SCYLV superpopulation structure, a second and also highly likely alternative leads to the T6 associated quadrangle. This is in contrast to the typical type of phylogram, such as that in Fig. 1B. Utilizing the same alignment, but a different phylogenetic model, the L1 and G2 isolates might be interpreted from Fig. 1B as being more closely allied with the B1 and N6 isolates, whereas the network diagram in Fig. 2 suggests that when other highly likely alternatives are considered, this relationship is more ambiguous. When SCYLV isolates are analyzed with alignments generated with other members of the family *Luteoviridae*, as shown in Fig. 3, a similar relationship between isolates may be interpreted. In both Fig. 2 and Fig. 3, the C1 isolate is positioned at a node within a quadrangle from which the C3 and C4 isolates are generated. Based upon these data, an assumption that any one of the three Columbian isolates could represent the closest C-population founder representative to the SCYLV superpopulation is also a reasonable interpretation.

The implications of our observations of SCYLV intraspecies SPV. Our analysis of SPV within North, South, and Central American isolates of SCYLV has been based in part on data derived from an intrafamilial model of SPV within the *Luteoviridae* (36). A direct extrapolation of data from this intrafamilial model to the SCYLV-A genome is shown in Fig. 4A. Based on this intrafamilial model, primers for the PROREP, REPUTR, and CPRT amplicons were selected from corresponding regions of the SCYLV genome, which yielded amplicons by annealing to sites with a low capacity for sequence evolution and yet produced amplicons that could each be predicted to have different degrees of sequence diversity. Among the predictions that could be made from our original intrafamilial analysis (36) are that the CPRT-spanning region of the SCYLV genome should exhibit a low sequence diversity, the PROREP-spanning regions should exhibit moderate diversity, and the REPUTR-spanning region should exhibit a higher diversity than the other two regions. As shown by our intraspecies analysis of SPV shown in Fig. 4A, the predictability of this intrafamilial analysis-based model of SPV is supported by the SPV that we detect from the SCYLV field isolates we have studied. The CPRT region exhibits low diversity, the PROREP region shows moderate diversity, and the REPUTR region exhibits high diversity. SPV models of this type have utilitarian value. They may be useful as criteria involved in the production of virus-resistant plants, utilizing transgenic methodologies that employ homology-dependent posttranscriptional gene-silencing mechanisms as a primary component. Because this method of plant protection is reliant upon transgene sequence homology with the targeted viral genome (2), the capacity for viral sequence evolution in the region represented by a transgene is a significant factor in developing ecologically safe and economically viable long-term virus resistance (34, 42, 49). For example, sugarcane plants expressing untranslated viral capsid sequences of *Sorghum mosaic virus* (SrMV) strain SCH (SrMV-SCH), challenged with SrMV viruses of strains SCM (SrMV-SCM) and SCI (SrMV-SCI) and *Sugarcane mosaic virus* strain D (SCMV-D), show various levels of virus resistance that correlated with the percentage of sequence identity of the transgenes to the sequence of the challenging virus (23). The corresponding homologous sequences of SrMV-SCM, SrMV-SCI, and SCMV-D, respectively, have 95, 95, and 75% identity with the equivalent SrMV-SCH sequence range (56). Challenge experiments with these same viruses protected, respectively, 17, 18, and 3 of 25 sugarcane plants compared to 22 of 25 plants protected by the SrMV-SCH virus-challenged plants (23). In many instances, sequence diversity data for a virus species is unavailable, but genomic information from related viruses is available. In these instances, an intrafamilial model could be constructed to predict genomic sequence regions expected to have low diversity, which would be more appropriate choices to use as transgene sequences.

ACKNOWLEDGMENTS

We thank James Miller (USDA-ARS Sugarcane Breeding Unit, Canal Pt., Fla.) for providing the F1 isolate source material used by Smith et al. (48) and James Irvine (Sugarcane Breeding Unit, Texas A&M Agricultural Experiment Station) for assistance in isolate collections during his visits to South and Central American locales and for helpful discussions on sugarcane biology and yellow leaf syndrome. We thank Tom Sappington (USDA-ARS Subtropical Agricultural Research Center, Weslaco, Tex.) for helpful discussions on population genetic analyses with molecular data. We also thank Grant Smith (Bureau of Sugar Experiment Stations, Indooroopilly, Queensland, Australia) and Mike Irey (U.S. Sugar Corp., Clewiston, Fla.) for helpful discussions on SCYLV and YLS.

This work was supported by grant 99-15 from the International Consortium of Sugarcane Biotechnologists.

REFERENCES

1. Bandelt, H. J., and A. W. M. Dress. 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Mol. Phylogenet. Evol. 1:242–252.
2. Baulcombe, D. C. 1996. Mechanisms of pathogen derived resistance to viruses in transgenic plants. Plant Cell 8:1833–1844.
3. Borth, W., J. S. Hu, and S. Schenk. 1994. Double-stranded RNA associated with sugarcane yellow leaf syndrome. Sugar Cane 3:5–8.
4. Chaloub, B. A., and H. D. Lapierre. 1996. Importance des recombinaisons ARN dans l'evolution des luteovirus. Agronomie 15:393–400.
5. Comstock, J. C., M. S. Irey, B. E. L. Lockhart, and Z. K. Wang. 1998. Incidence of yellow leaf syndrome in CP cultivars based on polymerase chain reaction and serological techniques. Sugar Cane 4:21–24.
6. Cronje, C. P. R., A. M. Tymon, P. Jones, and R. A. Bailey. 1998. Association of a phytoplasma with yellow leaf syndrome of sugarcane in South Africa. Ann. Appl. Biol. 133:177–186.
7. D'Arcy, C. J., and M. Mayo. 1996. Proposals for changes in luteovirus taxonomy and nomenclature. J. Gen. Virol. 142:1285–1287.
8. Domingo, E., C. Escarmis, N. Sevilla, A. Moya, F. Santiago, J. Q. Elena, I. S. Novella, and J. J. Holland. 1996. Basic concepts in RNA virus evolution. FASEB J. 10:859–864.
9. Domingo, E., and J. J. Holland. 1994. Mutation rates and rapid evolution of

RNA viruses, p. 161–189. *In* S. S. Morse (ed.), The evolutionary biology of viruses. Raven Press, New York, N.Y.
10. Domingo, E., J. J. Holland, C. Briebricher, and M. Eigen. 1995. Quasispecies: the concept and the word, p. 181–191. *In* A. J. Gibbs, C. H. Calisher, and F. Garcia-Arenal (ed.), Molecular basis of viral evolution. Cambridge University Press, Cambridge, England.
11. Domingo, E., and J. J. Holland. 1997. RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. 51:151–178.
12. Dopazo, J., A. W. M. Dress, and A. von Haeseler. 1993. Split decomposition: a new technique to analyze viral evolution. Proc. Natl. Acad. Sci. USA 90:10320–10324.
13. Dress, A. W. M., D. H. Huson, and V. Moulton. 1996. Analyzing and visualizing sequence and distance data using Splitstree. Discrete Appl. Math. 71:95–109.
14. Feselstein, J. 1992. Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genet. Res. 59:139–147.
15. Feselstein, J. 1989. PHYLIP: phylogeny interference package. Cladistics 5:164–166.
16. Gibbs, A. 1999. Evolution and origins of tobamoviruses. Proc. Philos. Trans. Soc. Lond. B 354:517–685.
17. Gibbs, M. J., and J. I. Cooper. 1995. A recombinational event in the history of luteoviruses, probably induced by base-pairing between two distinct viruses. Virology 206:1129–1132.
18. Grassly, N. C., and E. C. Holmes. 1997. A likelihood method for the detection of selection and recombination using sequence data. Mol. Biol. Evol. 14:239–247.
19. Hartl, D. L., and A. G. Clark. 1989. Principles of population genetics. Sinauer Associates, Sunderland, Mass.
20. Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.
21. Holmes, E. C., M. Worobey, and A. Rambaut. 1999. Phylogenetic evidence for recombination in dengue virus. Mol. Biol. Evol. 16:405–409.
22. Huson, D. H. 1998. Splitstree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73.
23. Ingelbrecht, I., J. E. Irvine, and T. E. Mirkov. 1999. Posttranscriptional gene silencing in transgenic sugarcane. Dissection of homology-dependent virus resistance in a monocot that has a complex polyploid genome. Plant Physiol. 119:1187–1197.
24. Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England.
25. Koonin, E. V., and V. V. Dolja. 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. 28:375–430.
26. Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: molecular evolutionary genetic analysis. Pennsylvania State University Press, University Park.
27. Lockhart, P. J., M. A. Steel, M. D. Hendy, and D. P. Penny. 1994. Recovering an evolutionary tree under a more realistic model of sequence evolution. Mol. Biol. Evol. 11:605–612.
28. Lockhart, P. J., D. Penny, and A. Meyer. 1995. Testing the phylogeny of swordtail fishes using split decomposition and spectral analysis. Mol. Evol. 41:666–674.
29. Maia, I. G., M. C. Goncalves, P. Arruda, and J. Vega. 2000. Molecular evidence that *Sugarcane yellow leaf virus* (ScYLV) is a member of the *Luteoviridae*. Arch. Virol. 145:1009–1019.
30. Martin, D. P., and E. P. Rybicki. 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16:562–563.
31. Martin, R. R., P. K. Keese, M. J. Young, P. M. Waterhouse, and W. L. Gerlach. 1990. Evolution and molecular biology of luteoviruses. Annu. Rev. Phytopathol. 28:341–363.
32. Mayo, M. A., and V. Ziegler-Graff. 1996. Molecular biology of luteoviruses. Adv. Virus Res. 46:416–468.
33. McLenachan, P. A., P. J. Lockhart, H. R. Faber, and B. C. Mansfield. 1996. Evolutionary analysis of the multigene pregnancy specific β1-glycoprotein family: separation of historical and nonhistorical signals. J. Mol. Evol. 42: 273–280.
34. Miller, W. A., G. Koev, and B. R. Mohan. 1997. Are there risks associated with transgenic resistance to luteoviruses? Plant Dis. 81:700–710.
35. Miller, W. A., S. P. Dinesh-Kumar, and C. P. Paul. 1995. Luteovirus gene expression. Crit. Rev. Plant Sci. 14:179–211.
36. Moonan, F., J. Molina, and T. E. Mirkov. 2000. *Sugarcane yellow leaf virus*: an emerging virus that has evolved by recombination between luteoviral and poleroviral ancestors. Virology 269:156–171.
37. Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70:3321–3323.
38. Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590.
39. Nei, M. 1972. Genetic distance between populations. Am. Nat. 106:283–292.
40. Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, Oxford, England.
41. Page, D. M., and E. C. Holmes. 1998. Molecular evolution: a phylogenetic approach. Blackwell Science, Cambridge, England.
42. Patron, N. J., M. A. Mayo, H. Barker, M. S. Liney, and H. G. Smith. 1998. Assessments of risk associated with growing virus resistant transgenic sugarbeet. Aspects Appl. Biol. 52:287–293.
43. Rathjen, J. P., L. E. Karageorgos, N. Habili, P. M. Waterhouse, and R. H. Symons. 1994. Soybean dwarf luteovirus contains the third variant genome type in the luteovirus group. Virology 198:671–679.
44. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
45. Scagliusi, S. M., and B. E. L. Lockhart. 2000. Transmission, characterization, and serology of a *Luteovirus* associated with yellow leaf syndrome of sugarcane. Phytopathology 90:120–124.
46. Schenck, S., and A. T. Lehrer. 2000. Factors affecting the transmission and spread of Sugarcane yellow leaf virus. Plant Dis. 84:1085–1088.
47. Schenck, S., J. S. Hu, and B. E. Lockhart. L. 1997. Use of a tissue immunoassay to determine the distribution of sugarcane yellow leaf virus in Hawaii. Sugar Cane 4:5–8.
48. Smith, G. R., Z. Borg, B. E. L. Lockhart, K. S. Braithwaite, and M. J. Gibbs. 2000. *Sugarcane yellow leaf virus*: a novel member of the *Luteoviridae* that probably arose by interspecies recombination. J. Gen. Virol. 81:1865–1869.
49. Smith, H., and N. Patron. 2000. Risk assessment of virus resistant transgenic sugar beet. Sugar Beet Rev. 68:47–50.
50. Strimmer, K., and A. von Haesler. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13:964–969.
51. Swofford, D. L. 1999. PAUP, version 4.04b. Sinauer Associates, Sunderland, Mass.
52. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity for progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Res. 22:673–680.
53. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24:4876–4882.
54. van Regenmortel, M. H. V., C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and R. B. Wickner. 2000. Virus taxonomy. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, New York, N.Y.
55. Vega, J., S. M. Scagliusi, and E. C. Ulian. 1997. Sugarcane yellow leaf disease in Brazil: evidence for associating with a luteovirus. Plant Dis. 81:21–26.
56. Yang, Z. N., and T. E. Mirkov. 1997. Sequence and relationships of *Sugarcane mosaic virus* and *Sorghum mosaic virus* strains and development of RT-PCR-based RFLPs for strain discrimination. Phytopathology 87:932–939.
57. Yeh, F. C., and T. J. B. Boyle. 1997. POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotech Centre, University of Alberta, Edmonton, Alberta, Canada.